PROBABILISTIC ESTIMATION OF PRECIPITATION

COMBINING GEOSTATIONARY AND TRMM SATELLITE DATA

Final Report

NASA Project NRA-02-ES-05 entitled

"**The Impact of Precipitation Measurement Missions on Hydrologic and Water Resource Predictions**"

Developed by

Carlo De Marchi
Doctoral Student, Georgia Water Resources Institute

and

Aris P. Georgakakos
Professor and Director, Georgia Water Resources Institute

Georgia Institute of Technology
Atlanta, Georgia 30332-0355

January 2008

ACKNOWLEDGEMENTS

The research described herein was funded in part by NASA (Goddard Space Flight Center, GSFC) under Project NRA-02-ES-05 entitled "The Impact of Precipitation Measurement Missions on Hydrologic and Water Resource Predictions." Additional funding was made available by other Georgia Water Resources Institute projects including the Nile Decision Support Program (Food and Agriculture Organization of the United Nations) and the Georgia Water Resources Program (United States Geological Survey). The authors wish to thank Dr. Christa Peters-Liddard for her technical support, guidance, and collaboration. We are also grateful to EUMESTAT for providing a large part of the METEOSAT data used in this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

1D	One-dimensional
2D	Two-dimensional
3D	Three dimensional
AVG_{nxn}	Average of variable over a region of nxn pixels
$STDEV_{nxn}$	Standard deviation of variable over a region of nxn pixels
AIVWNCC	Estimation algorithm based on analytical distributions, temporally and spatially correlated version
AIVWNCU	Estimation algorithm based on analytical distributions, temporally correlated, but spatially uncorrelated version
AIVWNUC	Estimation algorithm based on analytical distributions, temporally uncorrelated, but spatially correlated version
AIVWNUU	Estimation algorithm based on analytical distributions, temporally and spatially uncorrelated version
ANN	Artificial Neural Networks
AWES	Area-Weighted classification Error Score
B.U.R.S.L.	Bradford University Remote Sensing Limited
cdf	cumulative density function
ccdf	conditional cumulative density function
Dekad	Period of ten days

EIVWNCC	Estimation algorithm based on empirical distributions, temporally and spatially correlated version
EIVWNCU	Estimation algorithm based on empirical distributions, temporally correlated, but spatially uncorrelated version
EIVWNUC	Estimation algorithm based on empirical distributions, temporally uncorrelated, but spatially correlated version
EIVWNUU	Estimation algorithm based on empirical distributions, temporally and spatially uncorrelated version
Eumetsat	European Meteorological Satellite organization
FAO	Food and Agriculture Organization
FAR	False Alarm Rate
FFT	Fast Fourier Transform
Georgia Tech	Georgia Institute of Technology
GIS	Geographic Information System
GWRI	Georgia Water Resources Institute
IR	Thermal infrared radiation (~10 – 12 μm)
MAE	Mean Absolute Error
MFFNN	Multilayer Feed-Forward Neural Network
NileDST	Nile Decision Support Tool

NOAA	National Oceanic and Atmospheric Administration
PM	Passive microwave radiation (0.1 – 100 cm)
POD	Probability of detection
PR	TRMMPrecipitation Radar
SGSIM	Sequential Gaussian Simulation
SISIM	Sequential Indicator Simulation
SR	TRMM PR Surface Rain
TMI	TRMM Microwave Imager
TRMM	Tropical Rainfall Measurement Mission satellite
VIS	Visible radiation (~0.4 – 1.0 μm)
WMO	World Meteorological Organization
WV	Radiation in the 5.7 – 7.1 μm band, used for water vapor detection

SUMMARY

Environmental satellites represent an economic and easily accessible monitoring means for a plethora of environmental variables, the most important of which is arguably precipitation. While precipitation can also be measured by conventional rain gages and radar, in most world regions, satellites provide the only reliable and sustainable monitoring system. This report presents a methodology for estimating precipitation using information from the satellite-borne precipitation radar of the Tropical Rainfall Measurement Mission (TRMM). The methodology combines the precise, but infrequent, TRMM data with the infrared (IR) and visible (VIS) images continuously produced by geostationary satellites to provide precipitation estimates at a variety of temporal and spatial scales. The method is based on detecting IR patterns associated with convective storms and characterizing their evolution phases. Precipitation rates are then estimated for each phase based on IR, VIS, and terrain information. This approach improves the integration of TRMM precipitation rates and IR/VIS data by differentiating major storms from smaller events and noise, and by separating the distinct precipitation regimes associated with each storm phase. Further, the methodology explicitly quantifies the uncertainty of the precipitation estimates by computing their full probability distributions instead of just single "optimal" values. Temporal and spatial autocorrelation of precipitation are fully accounted for by using spatially optimal estimator methods (kriging), allowing to correctly assess precipitation uncertainty over different spatial and temporal scales. This approach is tested in the Lake Victoria basin over the period 1996-1998 against precipitation data from more than one hundred rain gages representing a variety of precipitation regimes. The precipitation estimates were shown to exhibit much lower bias and better correlation with ground data than commonly used methods.

Furthermore, the approach reliably reproduced the variability of precipitation over a range of temporal and spatial scales.

CHAPTER 1

INTRODUCTION

Each year, extreme events related to precipitation (i.e., floods and droughts), claim hundreds of lives and cause billions of dollars in damages. Not surprisingly, humanity has devoted large efforts to better measure and understand rainfall occurrence and distribution. Yet, reliable measurements of precipitation from the ground are restricted to few areas of the world, mostly in the more developed countries. Vast regions in the continents and most of the oceans are not routinely monitored, hampering the ability to adequately plan the use of the most important of life's necessities: water.

Satellite based remote sensing of precipitation has long been viewed as a promising tool for filling the gaps in the ground based precipitation measurements. In the last forty years, scientists have developed and improved an array of sensors and methodologies for estimating precipitation using satellite images. A major advance in remote sensing of precipitation is the deployment of an increasing number of satellites, such as the Tropical Rainfall Measurement Mission (TRMM), able to reliably measure rain rates, although at a low temporal frequency. The combination of this information with the more frequent data on cloud dynamics provided by geostationary satellites has been an intense focus of remote sensing research in recent years. The resulting estimation procedures have notably improved the capabilities of remote sensing of precipitation (e.g., Hsu et al., 1999; Adler et al., 2000; Joyce et al., 2004), but merging these two sources of information is still in the realm of active research.

Despite these enormous advances, however, remote sensing of precipitation is still affected by considerable uncertainty, even at coarse temporal and spatial scales (Adler et al, 2001).

Yet, most remote sensing estimates of precipitation provide no information on the estimation error. If information is provided, it is normally in the form of the estimation's mean square error at some spatial and temporal scale. However, this information is not very useful to users who need to aggregate precipitation over larger areas and/or periods or to assess the uncertainty in the hydrologic response of a basin to the estimated precipitation.

The objective of this research is to fill this gap in the available procedures for remote sensing of precipitation by exploring methodologies that produce reliable estimates of precipitation and quantify the associated uncertainty over any temporal and spatial scale of interest to the potential user.

Further, such methodologies must be based on information readily accessible in the underdeveloped areas of the world, where major is the need for such technologies. To pursue these objectives, this research developed a methodology to estimate precipitation using information from the satellite-borne precipitation radar of the Tropical Rainfall Measurement Mission (TRMM). The methodology combines the precise, but infrequent, TRMM data with the infrared (IR) and visible (VIS) images continuously produced by geostationary satellites (such as GOES and METEOSAT) to provide precipitation estimates at a variety of temporal and spatial scales. In contrast to most other merging techniques, the combination of the TRMM and geostationary data does not produce a single 'optimal' value but a full ensemble of equally probable values that can be used to assess the uncertainty in the precipitation estimate.

Further, to reduce the precipitation uncertainty, this research uses neural networks to recognize IR patterns associated with convective storms and their evolution phases. Precipitation rates are then estimated for each phase based on IR, VIS, and terrain information.

The research is organized in nine chapters. Chapter 2 gives some background on the precipitation phenomena and on remote sensing, and reviews past work on remote

sensing of precipitation. Chapter 3 outlines the conceptual framework of the probabilistic estimation methodology developed in the research and the steps followed to develop and test it. Chapter 4 describes the satellite images and gage data used to develop and validate the estimation procedures. Chapter 5 analyzes the methodology developed in this research to separate major convective storms from smaller events and noise, and to identify stages in the storm evolution. Chapter 6 discusses the precipitation distributions used in the estimation procedure and their relation with IR, storm stages, and orography. Chapter 7 describes the estimation procedure in detail and evaluates its performance in estimating precipitation over isolated spatial elemental units. Care is devoted to build the estimate ensembles in a way that properly represents the estimate uncertainty for any temporal duration of the estimate. Chapter 8 analyzes the procedures used to guarantee that estimate uncertainty is correctly represented when single-pixel precipitation is aggregated over larger areas. Chapter 9 discusses the overall performance of the developed methodology and addresses various aspects that can be improved further.

CHAPTER 2

REMOTE SENSING OF PRECIPITATION: BACKGROUND AND LITERATURE REVIEW

All water enters the land phase of the hydrologic cycle as precipitation. Assessing the amount and intensity of precipitation falling over an area during a certain period is, therefore, of fundamental importance for most branches of water resources management and for many other human activities such as agriculture, city planning, tourism, etc. The spatial and temporal resolution of the quantification of precipitation depends on the application of this information. Typically, agricultural planning and climatic study require monthly precipitation over areas of several thousands of square kilometers, although precipitation intensity during storms may also be necessary for evaluating the soil erosivity. This resolution must increase to 10-day and few thousands of square kilometers for irrigation and hydropower scheduling. Flood forecasting is typically interested in sub-daily or hourly precipitation at a resolution of several hundreds of square kilometers; although the recent trend towards the use of distributed hydrologic models require finer spatial resolutions. Finally, flash-flood warning and erosion assessment may require precipitation measurements at sub-hourly resolution over areas of few tens of square kilometers. Different methods for measuring precipitation – rain gages, precipitation radar, and satellites – can deal with this variety of spatial and temporal resolutions, with different degrees of success and at different costs, according to the precipitation characteristics and orography of the area of interest.

This chapter begins with a brief review of some characteristics of precipitation that are important for its measurement. The following section reviews the most common techniques for measuring precipitation using rain gages and precipitation radar. Section

2.3 first reviews the platforms and sensors that are presently available for remote sensing of precipitation. A special emphasis is given to the Meteosat and TRMM satellites that have been used for the development of the estimation technique presented in this study. The section continues examining the different techniques that have been used for remote sensing of precipitation using satellite images. Section 2.4 summarizes the most important aspects of remote sensing of precipitation.

2.1 Precipitation Phenomena

Practically all basic hydrology books briefly describe the fundamentals of precipitation formation (for example, Dingman, 2002). A more detailed treatment of the subject can be found in specialized books like Houze (1993). This section aims to highlight some aspects of precipitation and storms that are useful for better understanding the characteristics of the estimation procedure.

The formation of precipitation can be subdivided into four phases:

1. Moist air is cooled below the dew point to reach supersaturation;
2. Water vapor condenses into stable droplets/ice crystals;
3. Droplets or crystals grow into raindrops, snowflakes, or hailstones to sizes sufficient to surmount the updraft and survive evaporation during the descent;
4. Water vapor is continuously supplied to sustain the process.

Moist air cools below the dew point at rates sufficient to generate significant precipitation only when it is lifted to considerable elevations. During its ascension, moist air expands adiabatically, cooling at a rate of about 0.5 °C/100 m. The lifting of the moist air can take four general forms:

Frontal or *cyclonic convergence* – Masses of air move along horizontal gradients of pressure towards areas of low pressure. At the point of convergence of these flows, air

masses coming from opposite directions collide and are forced to rise. In the mid latitudes, masses of warm and moist air are displaced by masses of cold and dry air generating the cyclonic fronts.

Non-frontal or *horizontal convergence* – In the tropical belt the lifting is generated by the convergence of masses of warm and moist air.

Orographic –Masses of air are pushed by horizontal winds against a barrier, normally a mountain range, and forced to rise.

Convective – masses of air are heated at the Earth surface, becoming lighter than the surrounding atmosphere. Buoyancy force lifts the parcels.

Normally two or more of these lifting mechanisms interact during the lifetime of storms.

2.1.1 Stratiform and Convective Precipitation

Generally precipitation is called *convective* when updrafts are larger than one m/s. The strong lifting force carries water particles from the cloud base to high elevations through a moist environment, allowing them to enlarge substantially by accretion of cloud liquid water (Figure 2.1, times t_0, t_1, and t_2). Eventually, water droplets become large enough to overcome the updraft and begin to fall, continuing to increase in size by accretion during the descent (Figure 2.1, times t_3, t_4, and t_5).

When the updraft is less than one m/s, precipitation is called *stratiform*. This weak updraft of humid air maintains supersaturation, but is not able to suspend water particles in the upper atmosphere. Ice and water particles then begin to slowly fall through a moist environment, growing by rimming and aggregation. Depending on the temperature conditions, the precipitation may reach the ground as snow or it may melt while still in the atmosphere. In the latter case, water particles tend to aggregate and increase in size, falling more quickly (Houze, 1981, 1997).

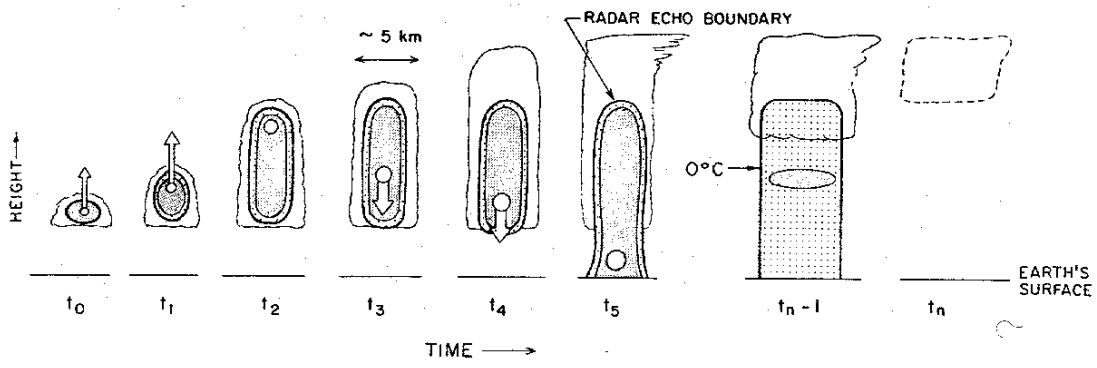


Figure 2.1 Generation of convective rain. (After Houze, 1981.)

Typically, convective precipitation presents radar vertical cross sections with the appearance of tall and thin columns of high reflectivity because large water and ice particles are distributed all along the vertical extension of the cell as in Figure 2.1 (Houze, 1997, 1993). In the horizontal plane, on the other hand, they look like isolated patches of high reflectivity at several altitudes (Steiner et al., 1995).

Stratiform rain is horizontally quite homogeneous showing reflectivity arranged in horizontal layers in radar cross sections (Figure 2.2). In particular, the melting layer appears as a horizontal band of high reflectivity, called the "bright band".

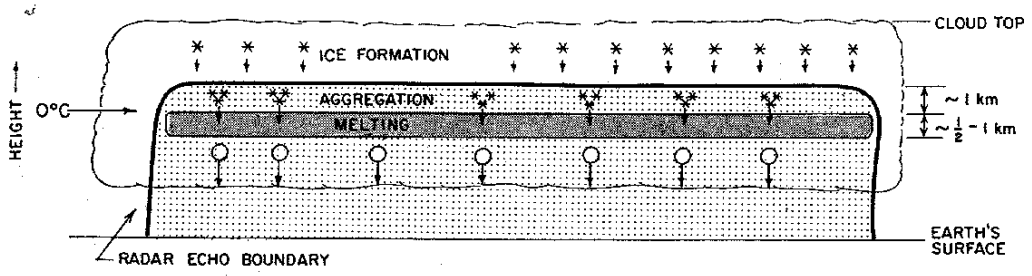


Figure 2.2 Generation of stratiform precipitation. (After Houze, 1981.)

Stratiform precipitation may be associated with either shallow or deep convection. In the first case, weak convective cells form in the upper layer of *nimbostratus* clouds, where potentially unstable air is present. The cells lift ice crystals through a moist environment, allowing them to grow. The ice crystals rapidly become too heavy for the weak updraft and fall.

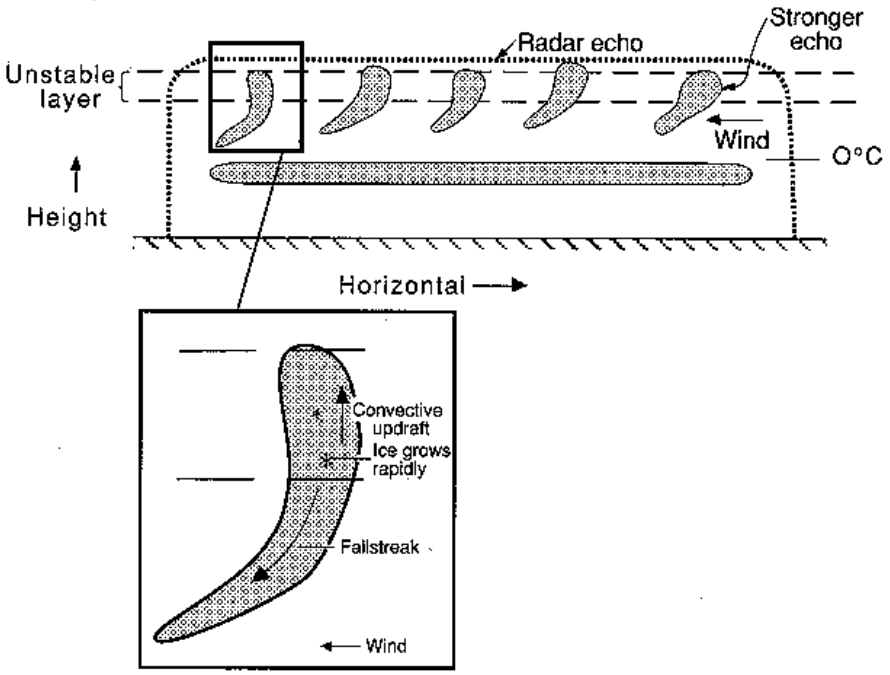


Figure 2.3 Stratiform rain generated by shallow convection. (After Houze, 1993.)

Deep convective cells lift in the upper troposphere large quantities of moisture, which aggregate into ice particles. When the convective updraft weakens, these particles slowly fall through a moist environment generating stratiform rain (Figure 2.4).

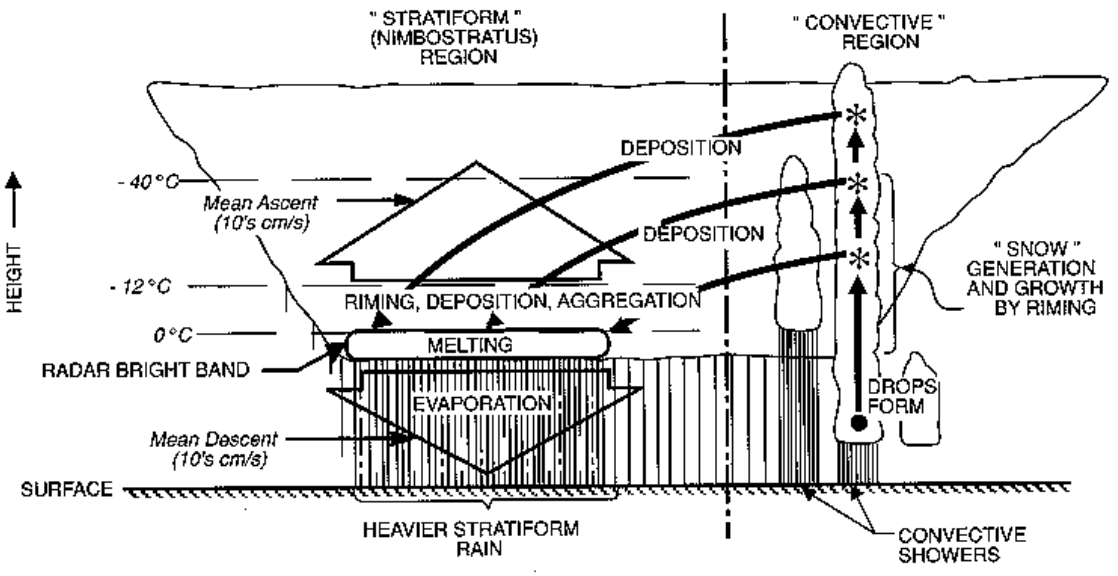


Figure 2.4 Stratiform rain generated by deep convection. (After Houze, 1993.)

A continuous chain of creation-extinction of convective cells characterizes the life of a storm and creates an extensive nimbostratus wake following the convective edge of the storm. The wake is not uniform, since the extinguishing convective cells preserve a stronger updraft than the older parts of the nimbostratus (Houze, 1993).

Convective rain is less frequent than stratiform rain, but because of its intensity (> 10 mm h^{-1}), convective rain accounts for around 50-70% of the precipitation volume in tropical storms (Houze, 1981; Churchill and Houze, 1984; Houze and Rappaport, 1984).

Stratiform rain is mild (<10 mm h^{-1}, but mostly ~ 1-3 mm h^{-1}), but it is more persistent and involves larger areas than convective rain. Consequently, its contribution is 30-50% of the total rain volume in the tropics and even more at mid latitudes.

2.1.2 Characteristics of Lifting Processes

The process responsible for lifting and cooling the moist air strongly influences the amount and intensity of the generated precipitation.

The *frontal systems* typical of the mid-latitudes are generated by the collision of masses of warm and moist air with masses of cold and dry air of polar origin, that develop cloud systems extending longitudinally for more than one thousand kilometers. On the eastern side of the front, warm air intrudes over stationary cold air, gently rising above it and generating the *warm front* (Figure 2.5). These fronts extend for up to 300-500 km in the east-west direction. On the western side of a front, cold and dry air moves eastward forcing warm and moist air upward generating the *cold front*. Typically, cold fronts are steeper than the warm fronts and generate clouds over a narrow band less than 80 km in width (Bras, 1990).

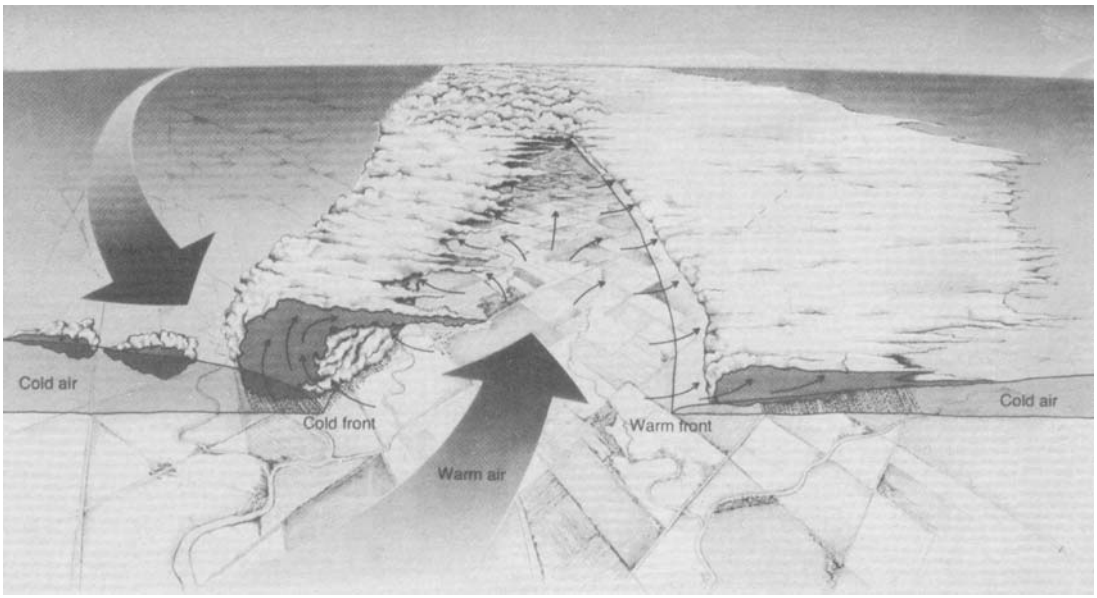


Figure 2.5 Transversal section of an extra-tropical cyclonic front. (After Marsh and Doziers, 1986.)

Though extratropical cyclones are typically associated with stratiform rain, the fronts are far from being uniform, presenting a complex texture of convective cells and wider areas of stratiform rain (Figure 2.6).

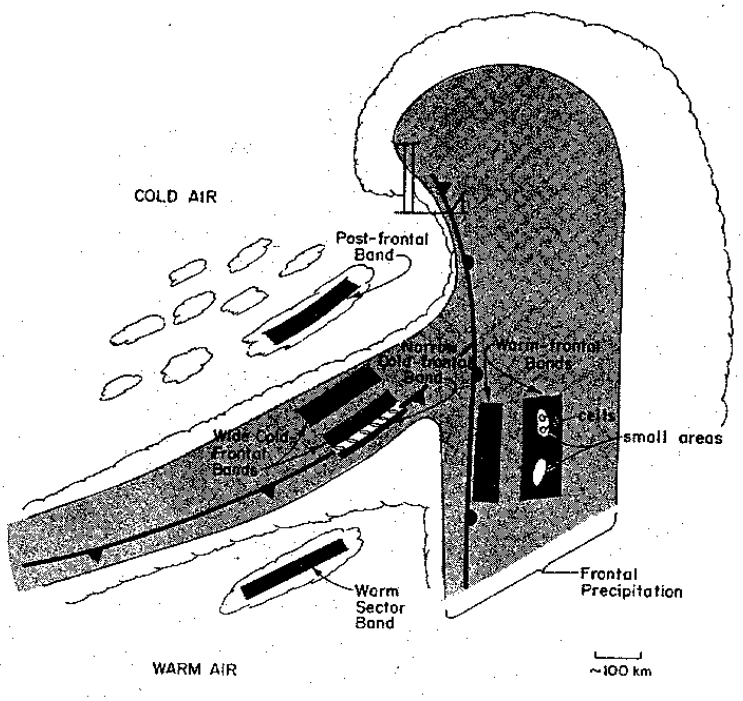


Figure 2.6 Spatial distribution of precipitation in extra-tropical cyclonic fronts. (After Houze, 1981.)

In *non-frontal* or *horizontal convergence*, the lifting is generated by the convergence of masses of warm and moist air. This type of lifting occurs mainly at the Inter Tropical Convergence Zone, where masses of warm air carried by the southern and northern components of the trade winds converge. Outside the tropics, horizontal convergence occurs occasionally at the interface of two adjacent low-pressure cyclones. Although the resulting precipitation is mostly stratiform (Houze, 1981), it may be intense and persistent, reaching up to 400 mm in 12 to 24 hours for tropical convergence (Bras, 1990).

Differential heating of the Earth surface creates differences in the temperature and, consequently, density in the air parcels in contact with it. Lighter parcels rise due to buoyancy forces, creating *convective* cells. These cells undergo a typical life cycle consisting of a *developing* or cumulus stage, a *mature* stage, and a *dissipating* stage

(Houze, 1981). During the first phase, clouds transform from cumuli into cumulonimbi due to rapidly rising warm, moist air (Figure 2.7). Rain droplets form at the cloud base and rise, carried by the strong updraft. During their upward trajectories, droplets grow by accretion of super-cooled water particles. If strong vertical shear is present, water particles are lifted well above the freezing level producing graupels and hailstones. Eventually, these particles become large enough to overcome the updraft and fall, dragging air with them and causing a downdraft. The beginning of precipitation indicates the onset of the mature phase of the storm, during which updraft and downdraft coexist. This phase typically lasts less than one hour (Dingman, 2002). When the updraft disappears, the cell enters the dissipating phase. During this phase previously lifted particles slowly fall, leading to lighter stratiform rain for long periods (Figure 2.4).

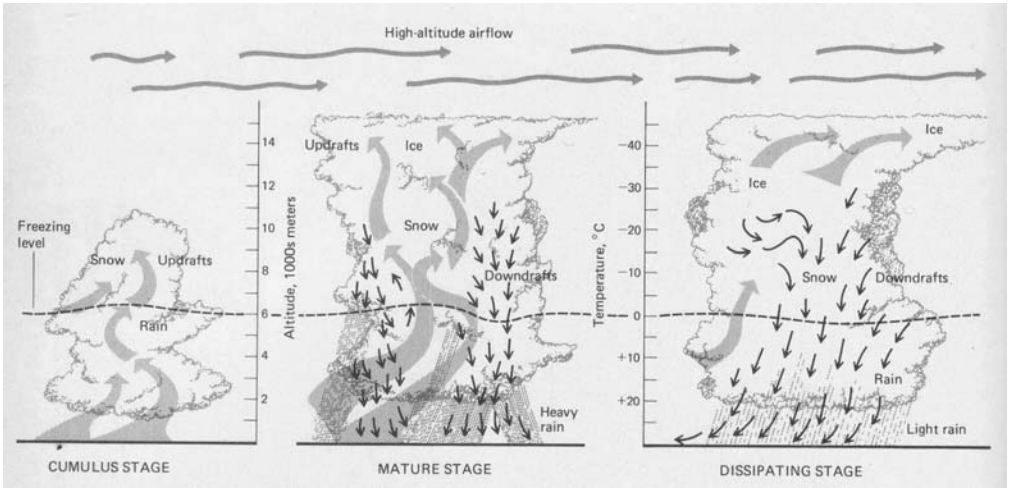


Figure 2.7 Section of a convective cell. (After Marsh, 1987.)

Mid-latitude summer thunderstorms are typical examples of convective storms. In the presence of weak wind shear, convective cells are almost randomly distributed in time and space within larger areas of stratiform rain and have a short lifetime. In stronger

vertical shear conditions, convective cells develop near the leading edge of the storm, giving rise to the so called "multicell" thunderstorm, which features heavier precipitation and, sometimes, hail. Occasionally, the organization of these cells is so tight that they behave as a single gigantic cell (Figure 2.8). This type of thunderstorm is called a "supercell" storm and is characterized by strong updraft with the consequent production of hailstones (Houze, 1981).

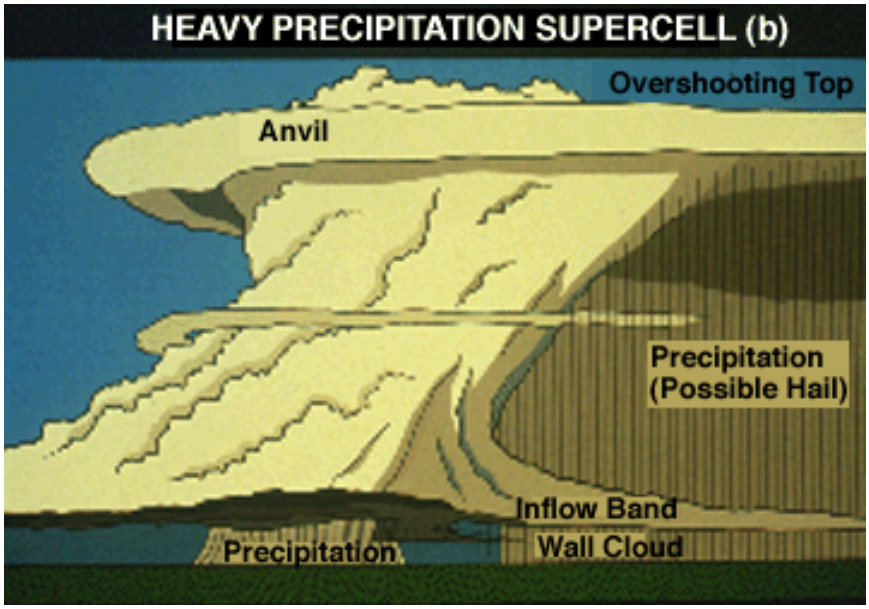


Figure 2.8 Super-cell thunderstorm. (After, University of Illinois, 2002.)

In the tropics, convective phenomena generate most precipitation events, from isolated small showers covering a few square kilometers to large mesoscale convective systems covering tens of thousands of square kilometers. While small showers account for the vast majority of events, the large cloud clusters produce most of rainfall. In fact, in the tropical Pacific, the largest 1% of storms may account for as much as 40% of the total rain (Houze, 1993).

The most commonly known *orographic* precipitation is generated when a mass of air is pushed by horizontal winds against a barrier, normally a mountain range, and forced to rise (Figure 2.9).

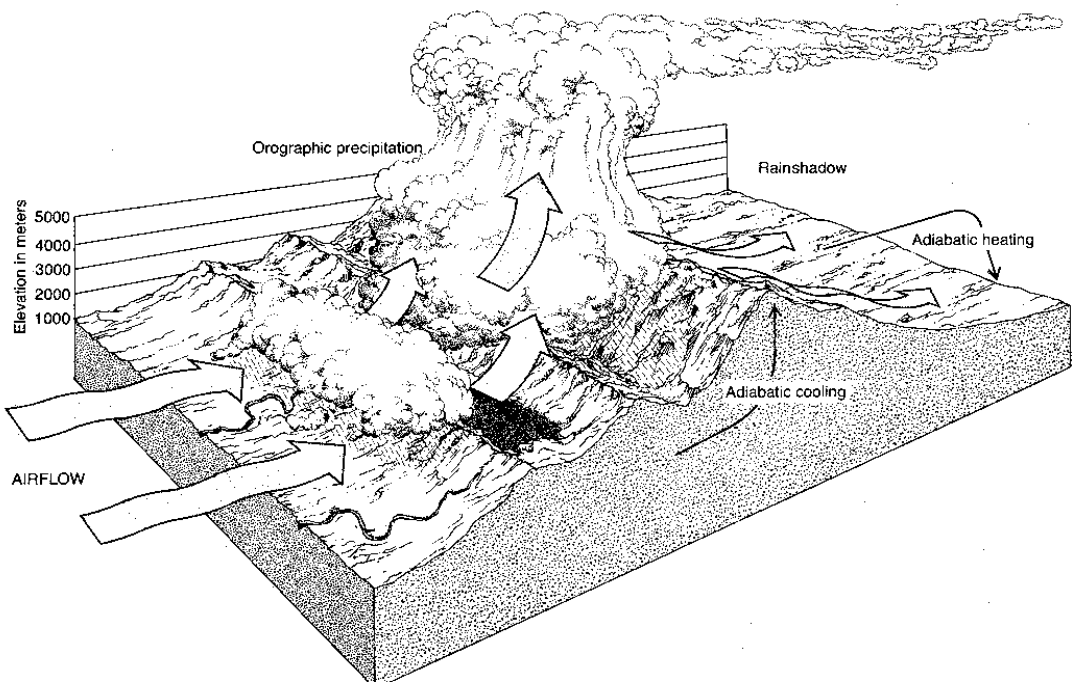


Figure 2.9 Main mechanism of orographic precipitation. (After Marsh and Doziers, 1986.)

Clouds and precipitation form on the windward side of the mountain range. Beyond the crest of the ridge, the air mass tends to subside and warm, stopping the generation of rain. For this reason, leeward sides of mountains are generally drier than the windward sides. In many situations, however, orography interacts in more complex ways with precipitation phenomena associated with other lifting mechanisms (Figure 2.10).

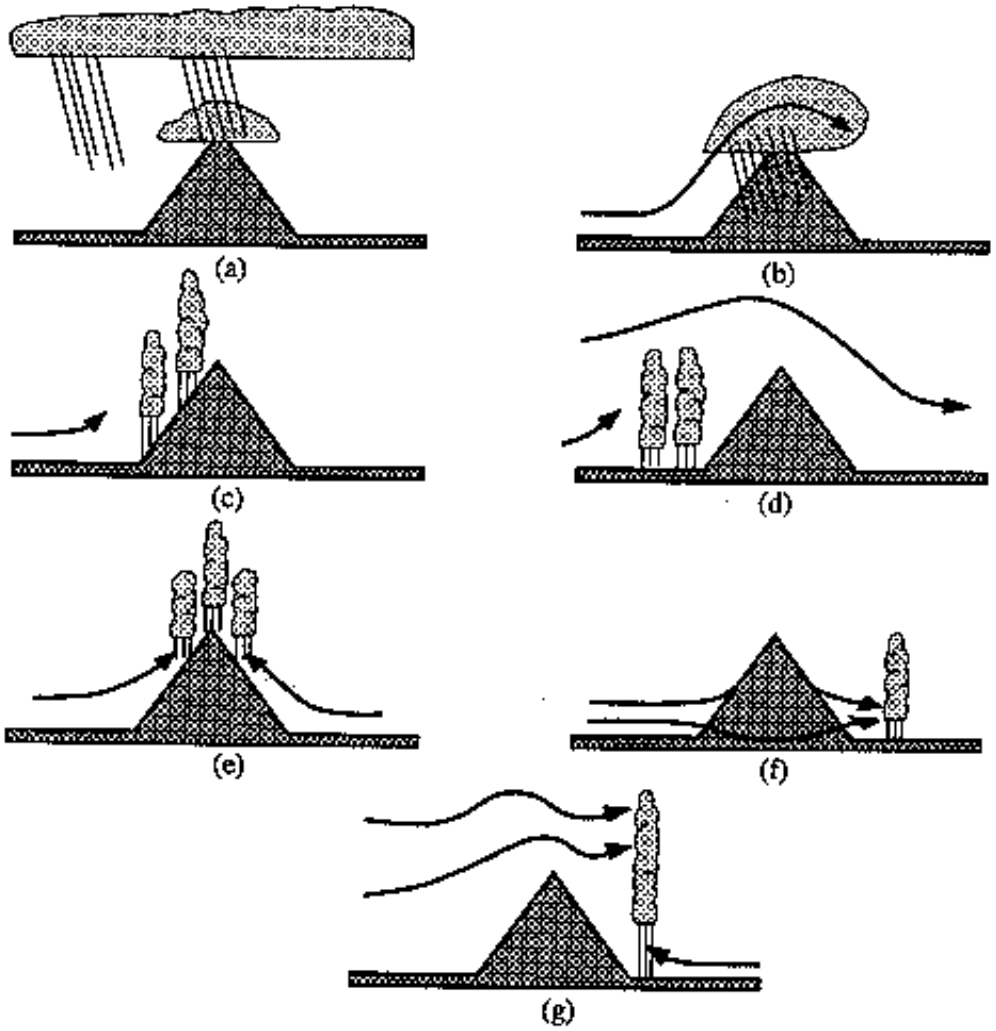


Figure 2.10 Orographic enhancement of precipitation. a) Seeder-feeder mechanism, b) Upslope condensation, c) Upslope triggering of convection, d) Upstream triggering of convection, e) Thermal triggering of convection, f) Lee-side triggering of convection, and g) Lee-side enhancement of convection. (After Houze, 1993.)

2.2 Precipitation Measurement from Ground

Measuring rainfall with conventional or remote sensing techniques is particularly challenging due to the high spatial and temporal variability of precipitation. Spatially, instantaneous precipitation intensity may vary from zero to more than 125 mm h^{-1} within a few kilometers. At the daily scale, this spatial variability is still very significant, while monthly and yearly patterns become more homogeneous. Even so, monthly and yearly

precipitation regimes may vary within a few kilometers due to orography, land/water interfaces, and large urban areas. Temporally, however, precipitation remains highly variable across a wide range of scales (i.e., sub-daily to seasonal and inter-annual resolutions).

Furthermore, rainfall intensity at fine temporal resolution is highly skewed towards zero, making its statistical description more difficult (Kidd, 2001). It is only at coarse resolutions that the distribution of precipitation tends toward normality.

2.2.1 Rain Gages

Rain gages are the oldest and most widely employed instrument for measuring precipitation. The precipitation depth during a particular event is determined by dividing the volume of rain collected in the measuring device by the area of the rain gage intake. Non-recording gages are merely graduated cylinders that are manually read at regular intervals, typically once per day. Although very simple in principle, the accuracy of these instruments depends on wind conditions, evaporation intensity, and dedication of the operator to read the gage always at the same time, even during intense storms.

Recording gages use weight recording, a tipping-bucket, or optical devices (Dingman, 2002) to measure precipitation at finer temporal resolutions and do not depend upon the punctuality of a human reader. However, there are additional concerns about the ability of these mechanisms to measure intense precipitation and the possibility of mechanical and electrical failures. Even in the best operative conditions, errors in rain gage data may reach 75% for individual events and 5-15% for long-term totals (Dingman, 2002).

Meteorological and water resources agencies usually provide precipitation data in the form of daily, monthly, or yearly summaries of many stations, a process prone to mistakes such as transcription errors and incorrect unit conversions.

For climatological studies and long-term planning, point measurements of precipitation are useful, but are of limited interest for many hydrological applications such as flood forecasting and hydropower scheduling. Such applications require estimates of the precipitation over an entire watershed, either as average (Mean Areal Precipitation) or as a distributed field. There are several ways to convert point measurements into MAP, ranging from simple techniques that do not require any knowledge of the local precipitation process (Thiessen polygons, inverse square distance, spline surface, etc) to complex algorithms involving extensive studies of the precipitation distribution in the region (optimal interpolation or kriging). Several authors have concluded that kriging-based algorithms provide the most accurate estimates of the spatial distribution of precipitation for most gage densities and temporal durations, although differences among the various methods are often less than 10% (Dingman, 2002). The success of these interpolation schemes in representing a precipitation field depends on the density of the rain gage network, topography of the area, and temporal duration. Table 2.1 reports the minimum density of rain gages recommended for general hydrological purposes. For quality control purposes, it is also advisable to have two different rain gages at each measurement point.

Historically, the creation of rain gage networks has been conditioned more by issues of convenience of access and cost, than by a desired accuracy in MAP estimation. This sometimes results in nominally dense networks that are not very useful for estimating precipitation over sparsely populated or topographically challenging areas. The advent of data loggers, wireless communication, and solar panels has improved the capability of deploying rain gages in remote areas. However, the relatively high value of these devices may limit their extensive use in some parts of the world.

Table 2.1 Minimum rain gage density recommended for various climatic situations. (WMO, 1981.)

Geographic Region	Km^2/gage
Small mountainous islands with irregular precipitation	140-300
Temperate, Mediterranean, and tropical mountainous regions	300-1000
Flat areas in temperate, Mediterranean, and tropical regions	1000-2500
Arid and polar regions	5000-20,000

2.2.2 Precipitation Radars

Although radars have been used for studying meteorological events practically since their invention in the 1940s, only in the last twenty years they have been used for the operational estimation of precipitation. The estimation of precipitation with radar is based on the principle that raindrops and snowflakes, as all objects, reflect part of the electromagnetic radiation reaching them. The power reflected (*Pr*) by a small segment V of the radar beam filled with raindrops or ice particles at a distance *r* from the radar is (Rinehart, 1991; Austin, 1987):

$$P_r = \frac{C|K|^2 z}{r^2}$$
$$K = \frac{m^2 - 1}{m^2 + 1} \qquad (2.1)$$
$$z = \sum D_i^{\,6}$$

where C = radar constant, a function of the radar characteristics (frequency, emitted power, antenna gain, beam width and pulse length)

m = index of reflection of water or ice;

r = distance (km);

z = radar reflectivity factor of volume V ($mm^6\ m^{-3}$);

D_i = diameter of raindrops or ice particles i (mm);

Equation 2.1 allows the computing of the reflectivity of the target from the measured reflected power and the knowledge of the type of target (raindrops or ice particles), but it gives only a theoretical relation with the precipitation rate. Determining this variable is possible by using experimental relations between rain rates and radar reflectivity in the form:

$$z=a*R^b \qquad (2.2)$$

where R= rain rate (mm h^{-1})

a, b = empirical parameters dependent on storm type and location

Several dozens of statistically optimal *a* and *b* couples have been determined for different locations and storms, but in practice three or four of these sets cover the range of possible precipitation patterns with sufficient accuracy (Rinehart, 1991). Commonly used values are a=200, b=1.6 for rain (Marshall and Palmer, 1948) and a=2000, b=2.0 for snow (Carlson and Marshall, 1972).

Radars have several advantages over rain gages:

- They provide the rain rate distribution over areas of tens of thousands square kilometers at a resolution of 1-4 km, a feat that would require thousands of gages.
- They can measure rain rate as an average rate over a 1-16 km^2 area instead of a point measurement.
- They provide a three dimensional picture of storms at a typical rate of an image every 5-10 minutes.
- They centralize the data collection procedure, making it easier to avoid accidental errors.

However, precipitation radars have also some notable drawbacks:

- Several environmental variables (true rain rate/reflectivity function, presence of hail, partial beam filling, radiation attenuation, clutter, beam elevation from the ground, anomalous propagation, etc.) affect their estimates causing quantitative errors of 50-100% frequently even at distances of less than 100 km (Dingman, 2002; Rinehart, 1991). However, the integration of radar data and even a relatively sparse network of rain gages may limit the radar error to 10-15% (Rinehart, 1991).
- Radars are relatively expensive to install and maintain. Further, to be used effectively, they require highly trained personnel. Additional hardware is required to integrate radar and rain gage data. These characteristics limit their use in countries were resources and adequately trained scientists are scarce.
- The use of ground radars may be severely limited in areas with high relief since mountains block the propagation of electromagnetic radiation,

2.3 Precipitation Measurement from Satellites

As section 2.2 pointed out, ground based measurement of precipitation has several shortcomings:

- Rainfall is characterized by high spatial and temporal variability, making its quantification using rain gages challenging, especially at high spatial and temporal resolution.
- Rain gages are mostly located in easily accessible areas. Unpopulated areas, especially mountains, are undermonitored, despite the fact that in these areas precipitation is often intense and frequent.

- Because rain data are often collected by different agencies, according to different procedures, it is difficult to organize them in common databases to cover large regions.
- Rain gages vary in shape and temporal resolution.
- The accuracy of rain gages depends on their design and environmental factors (wind speed and direction, temperature, humidity, etc).
- Rain gage maintenance and rain data quality control is costly and difficult.
- Few rain gages provide real time data, especially outside the more developed countries.
- Rainfall radars are an answer to some of these problems. However, they are costly, require maintenance, and introduce significant errors.

Satellites potentially solve several of these problems:

- Collection, storage, and distribution of data are centralized.
- Highly qualified personnel control the data according to standardized procedures.
- Spatial distribution of precipitation is available at the spatial resolution of 1-5 km and temporal resolution of 0.5-1 hour.
- Precipitation data can be made available within one hour to users through internet or satellite broadcasting.
- Most importantly, satellites are able to measure precipitation over undermonitored areas, including the oceans, allowing a better understanding of the global climate and its change.

For all these reasons, in the last fifty years, a lot of research has been devoted to developing platforms, sensors, and procedures to estimate precipitation from satellites.

Subsection 2.3.1 gives some basic notion about remote sensing that may help clarify the procedures for estimating precipitation from satellites. More detailed treatments of this subject can be found in Elachi (1997) and Kidder and Vonder Harr (1995). Sub-section 2.3.2 reviews the platforms and sensors available for the remote sensing of precipitation. Finally, sub-section 2.3.3 examines the techniques that have been developed for the remote sensing of precipitation.

2.3.1 Remote Sensing Principles

Although some sensors are based on gravitational and magnetic effects, observation of the Earth from satellites is based mostly on measuring the electromagnetic radiation emitted, reflected, and absorbed by the Earth and the atmosphere (Elachi, 1997). In this case, the remote sensing process can be broken down into seven components (Figure 2.11):

A) *Energy source or illumination* – In many remote sensing applications, either the sun or the sensor carried by the spacecraft supply electromagnetic energy to the target of interest. In other cases, the target itself emits the observed electromagnetic energy.
B) *Radiation-atmosphere interaction* - As the energy travels from its source to the target and from this to the sensor, it will interact with the atmospheric elements it passes through.
C) *Interaction with the target* – The interaction between the electromagnetic radiation illuminating the target and the target is a function of the characteristics of both the radiation and the target itself.
D) *Detection of energy by the sensor* – A sensor not in contact with the target measures the energy that has been scattered or emitted by the target in specific electromagnetic bands. Typically, large observed areas are subdivided into matrixes of smaller elements, called pixels. The sensor records the energy coming from each of them.

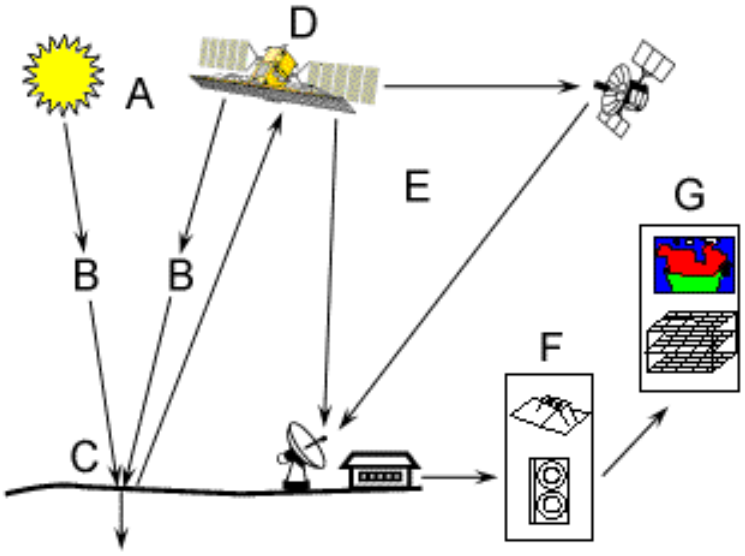


Figure 2.11 Components of a remote sensing systems. (After Canada Centre for Remote Sensing, 2002.)

E) *Transmission, reception, and processing* – The energy recorded by the sensor is quantized into numerical data, that are then encoded and transmitted, sometimes through a relay, to a receiving station. Here, the data are decoded and translated into digital images.

F) *Interpretation and Analysis* – The digital images are interpreted, visually and/or electronically, to extract information about the observed target.

G) *Application* – The final element of the remote sensing process is achieved when the information extracted from the imagery about the target is applied to solve a particular problem.

Remote sensing of precipitation usually exploits radiation in the visible (wavelengths between 0.4 to 0.7 μm), infrared (0.7 to 30 μm), and microwave (1 mm to 10 cm) bands (Figure 2.12).

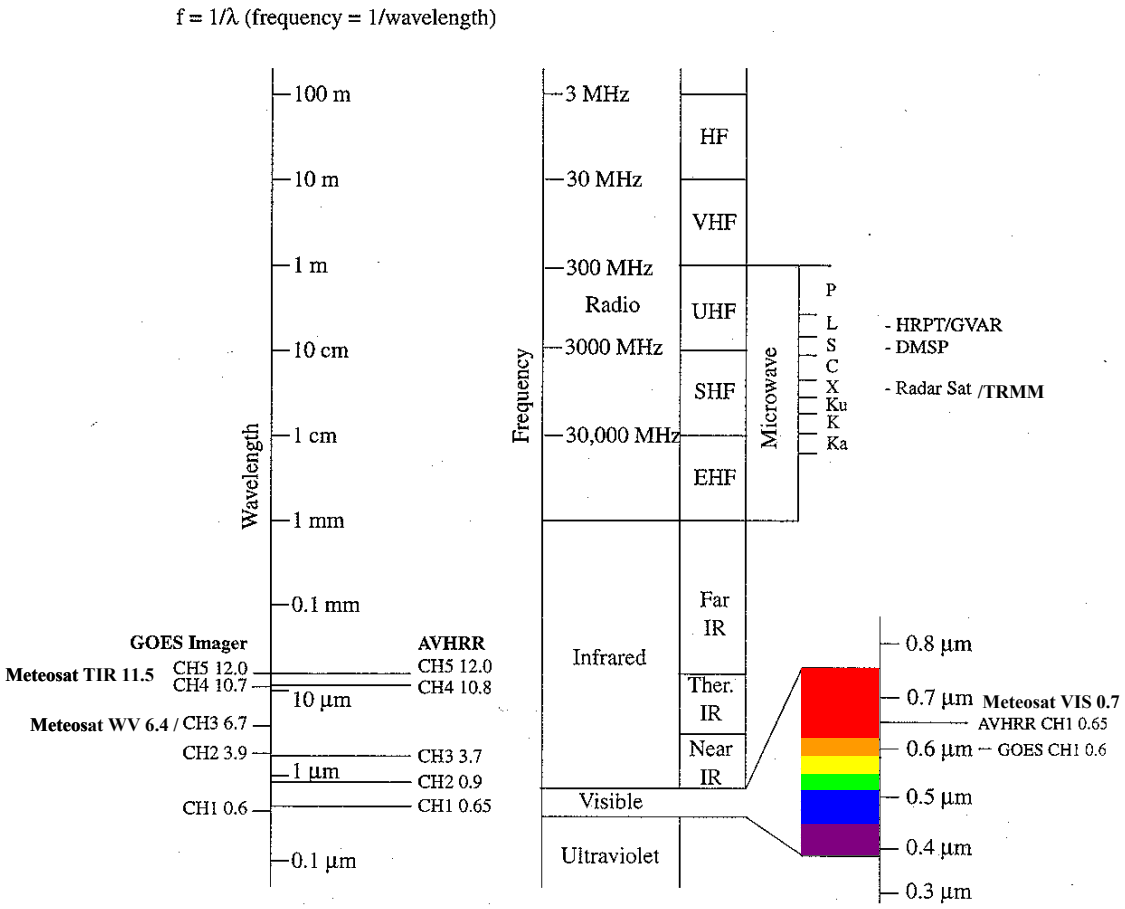


Figure 2.12 The electromagnetic spectrum. (After Seaspace, 1999.)

Molecules and atoms excited to high energy levels release electromagnetic radiation when they decay to lower energy states. The wavelength of this radiation is inversely proportional to the difference in the energy levels. Materials excited by electric discharges, nuclear decay, and chemical reactions are characterized by well-defined changes in energy levels, which produce radiations in very specific and narrow spectral bands. Heating, on the other hand, raises the molecule excitation (electronic, vibrational, and rotational) over a wide array of levels. The resulting radiation has a wavelength that decreases with the temperature of the material and is distributed over a broader band (Plank's law).

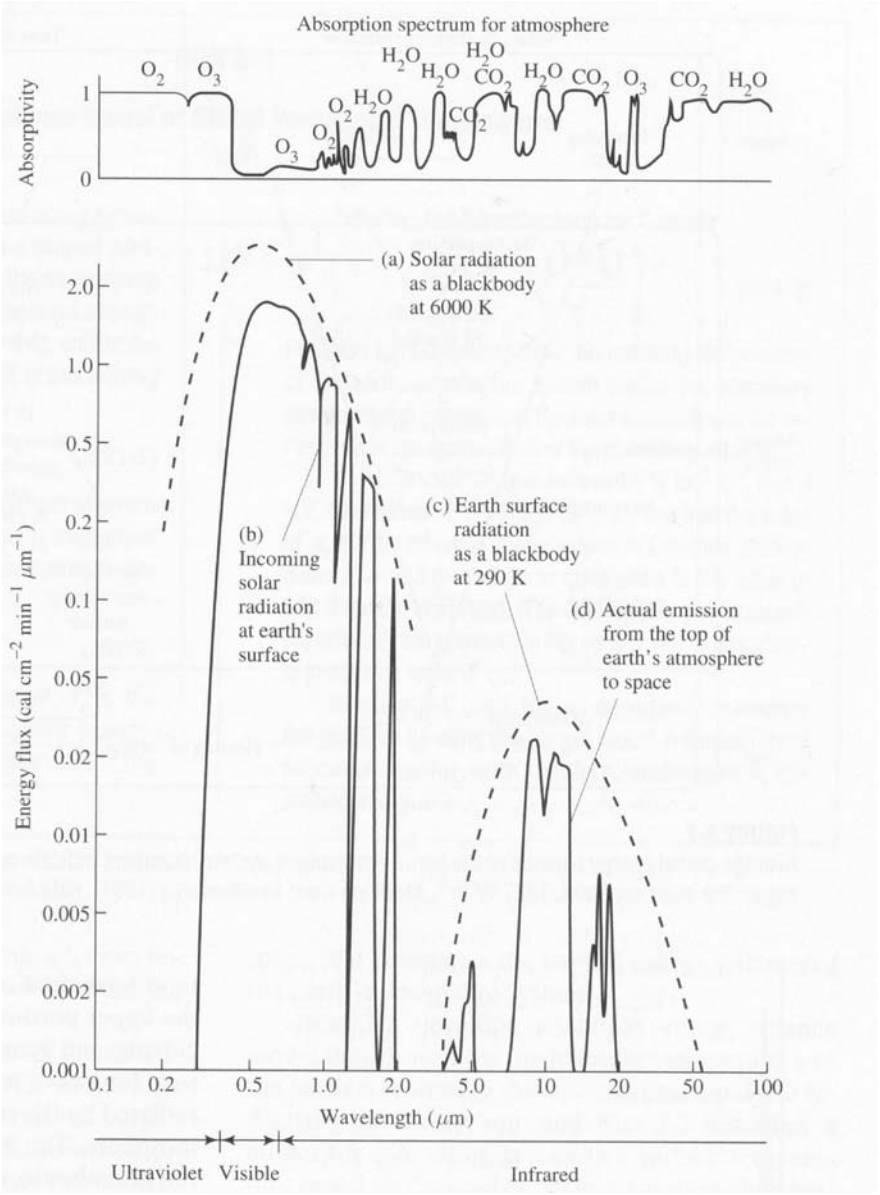


Figure 2.13 Spectra of energy (a) emitted by a blackbody at 6000 K, (b) received at the Earth's surface (global average), (c) emitted by a blackbody at 290 K, (d) emitted to space by the Earth-atmosphere system (global average). Upper graph shows absorption spectrum of the atmosphere. (After Dingman, 2002.)

The solar radiation reaching the top of the atmosphere is almost equivalent to that of a perfect emitter (blackbody) at 6000 °K (Figure 2.13). The atmosphere strongly absorbs the solar radiation in the ultraviolet band and in parts of the near-infrared band, but transmits almost completely the visible component. Clouds, oceans, and land

surfaces, on the other hand, reflect part of the incident solar radiation back to space. The fraction of the total incident visible radiation reflected by a certain material is called *albedo* and is a property of the material itself (Table 2.2). Given that the Earth and atmosphere do not emit visible radiation and that the atmosphere is relatively transparent in this band (Figure 2.13), the visible radiation detected by satellites can be related to the nature of the reflecting surfaces.

Remote sensing of precipitation takes advantage of the fact that the albedo of clouds is higher than that of oceans and most land surfaces, and that cloud albedo increases with cloud thickness (Table 2.2).

King et al. (1995) showed that precipitation rate correlates better with visible radiation than it does with infrared radiation. However, the usefulness of visible radiation for precipitation estimation is limited by the fact that it is available only during daytime and that most satellites do not have on-board devices for continuously recalibrating the relation between observed radiation and albedo. Further, clouds may be difficult to detect against light surfaces such as snow and sand (Isaacs, 1993).

Because of their lower temperature, Earth and its atmosphere emit radiation only in the infrared band, where the interference from reflected solar radiation is relevant only for wavelengths below 3 μm (Figure 2.13). The atmosphere is more or less opaque in some parts of the infrared band, but it is almost transparent in the 10-12 μm interval. Consequently, the radiation observed in this last interval is generated only by thermal excitation phenomena. and can be accurately translated into temperature maps of the Earth surface and overlaying clouds using Plank's formula. Further, in this region of the infrared band, thick clouds behave as perfect blackbodies, completely absorbing the radiation coming from the surface and lower atmosphere, and emitting upward radiation only from their upper layers (Kidder and Vonder Haar, 1995). Thus, the upward radiation emitted by thick clouds can be related to the temperature of the cloud's top, which in turn is related to the elevation of the cloud's top.

Table 2.2 Integrated visible-range reflectance (Albedo) of various surfaces. (After Dingman, 2002.)

Surface		Conditions	Albedo
Clouds	Low overcast:	100 m thick	0.40
		200 m thick	0.50
		500 m thick	0.70
		2000 m thick	0.80
Liquid water	Smooth	Solar angle 60°	0.05
		Solar angle 30°	0.10
		Solar angle 20°	0.15
		Solar angle 10°	0.35
		Solar angle 5°	0.60
	Wavy	Solar angle 60°	0.10
Solid water	Fresh snow	Low density	0.85
		High density	0.65
	Old snow	Clean	0.55
		Dirty	0.45
	Glacier ice	Clean	0.35
		Dirty	0.25
Sand	Dry, light	High sun	0.35
		Low sun	0.65
	Gray	Wet	0.10
		Dry	0.20
	White	Wet	0.25
		Dry	0.35
Soil	Organic	Dark	0.10
	Clay		0.20
	Sandy	Light	0.30
Grass	Typical fields		0.20
	Dead	Wet	0.20
		Dry	0.30
Tundra, heather			0.15
Crops	Cereals, tobacco		0.25
	Cotton, tomatoes, potatoes		0.20
	Sugar cane		0.15
Trees	Rain forest		0.15
	Eucalyptus		0.20
	Red pine forest		0.10
	Mixed deciduous hardwoods		0.18

Remote sensing of precipitation takes advantage of this relationship between thickness and temperature for detecting the presence of thick and tall clouds, which

appear as colder than the surrounding areas and are strongly related to precipitation. Satellites images in the infrared band have the major advantage of being available for the entire 24 hours, but they are not very useful when the precipitation is generated by low level clouds (warm rain, often of orographic origin), or when the rain bands are embedded in larger areas of non-raining high clouds (frontal systems).

Thermal radiation from the Earth and the atmosphere is stronger in the infrared region, but extends also into the submillimeter and microwave region (Figure 2.12), where Plank's law simplifies into a linear relation between emitted radiance and temperature of the emitting body. Remote sensing in this region of the spectrum is particularly attractive because, for large segments of the band, the atmosphere (Figure 2.14) and even non-precipitating clouds are almost transparent (Elachi, 1997).

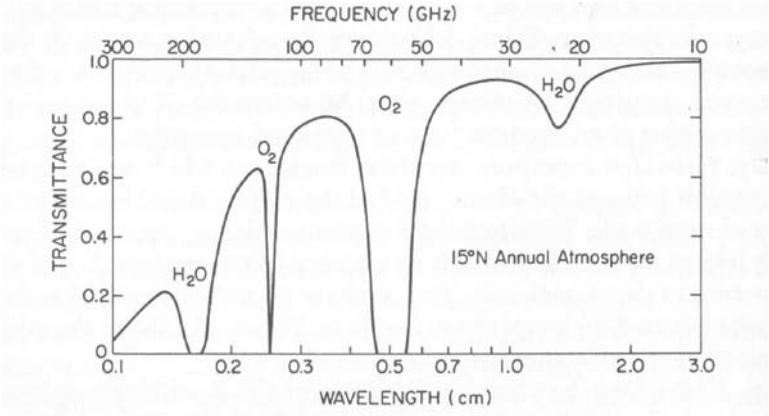


Figure 2.14 Transmittance of the Earth's atmosphere in the microwave region. (After Kidder and Vonder Haar, 1995.)

Unlike clouds, raindrops and ice particles strongly interact with microwave radiation, with absorption and emission dominating frequencies below 22 GHZ and scattering dominating frequencies above 60 GHZ (Kidder and Vonder Haar, 1995; Kidd

and Barrett, 1990). Remote sensing of precipitation takes advantage of both these effects for estimating precipitation.

At frequencies below 20 GHZ, water surfaces, such as oceans and lakes, behave as graybodies with emittance roughly equal to 0.5. Consequently, these surfaces appear to have only about half the real temperature of the surface, looking very "cold" to a passive microwave (PM) radiometer. The emittance of raindrops, on the other hand, is close to 0.9, producing a brightness temperature that equals the raindrops' temperature and offering a strong contrast against "cold" water surfaces. The more raindrops, the warmer the whole scene appears, making possible to obtain fairly accurate rainfall rates based on the temperature of the microwave scene. Land, on the other hand, behaves more similarly to a blackbody with emittance close to 0.9. Consequently, the thermal contrast introduced by the raindrops is reduced, making precipitation estimation more difficult.

Frequencies above 60 GHz (for example 85.5 GHz) are strongly scattered by the ice present in many raining clouds. Thus, clouds containing ice reduce the surface-generated microwave radiation able to reach the satellite sensors, offering a strong contrast against the warm background. Because land has higher blackbody temperatures and because the storm ice layers are normally thicker overland, these frequencies are exploited preferably for estimating precipitation over land.

Most algorithms for estimating instantaneous rainfall rates combine the radiance sensed at different frequencies, according to empirical or physically based theoretical models.

2.3.2 Platforms and Sensors for Remote Sensing of Precipitation

Satellites used for observing the Earth and the atmosphere can be subdivided into two basic categories according to the orbit where they are placed: geostationary satellites and low-Earth orbit (LEO) satellites. This subdivision is fundamental because the orbit

of the satellite determines the type of sensors it can carry and their spatial and temporal resolution.

2.3.2.1 Geostationary Satellites

Geostationary satellites fly in circular orbits on Earth's equatorial plane at an altitude of around 35,800 km above the mean sea level (Figure 2.15). Theoretically, at this distance the Earth's gravitational force perfectly balances the centrifugal force of satellites rotating at the Earth's angular velocity. The uneven shape of the planet and the gravitational influences of the sun and the moon, however, perturb the position of the satellites, requiring periodic maneuvers to keep them within a reasonable range from the nominal value. This operation is easily accomplished by operating the thrusters of the satellite, but it consumes the reserves of fuel available to the satellites, thus limiting their maximum operational life (Eumetsat, 2000; Kidder and Vonder Haar, 1985). Satellites orbiting in the Earth's equatorial plane and at the Earth's angular speed are stationary relative to a given position on the Earth's surface.

Geostationary satellites have the major advantage of being able to continuously monitor a given area of the planet, producing an image of the entire hemisphere every 15-30 minutes at a resolution of 1-5 km. On the other hand, due to their distance from the Earth, geostationary satellites can carry only a limited range of instruments. For precipitation estimation, this normally means a multi-channel radiometer, which senses reflected solar visible and infrared radiation and the infrared radiation emitted by the Earth and atmosphere. Further, geostationary satellites have a distorted view of the areas beyond 60° of the great circle arc from the sub-satellite point (i.e. locations outside the circle generated by the intersection of the Earth's surface and a cone with its axis on the line connecting the satellite to the center of the Earth, the vertex on the center of the Earth, and an aperture of 60°). This means that locations north or south of 60 degrees latitude cannot be properly observed. In addition, during the periods from March 1 to

April 15 and September 1 to October 15 of each year, geo-stationary satellites remain in the Earth's shadow for up to 70 minutes around midnight local time. During this time, the satellite's photovoltaic panels cannot generate power, forcing the satellite to turn off its sensors.

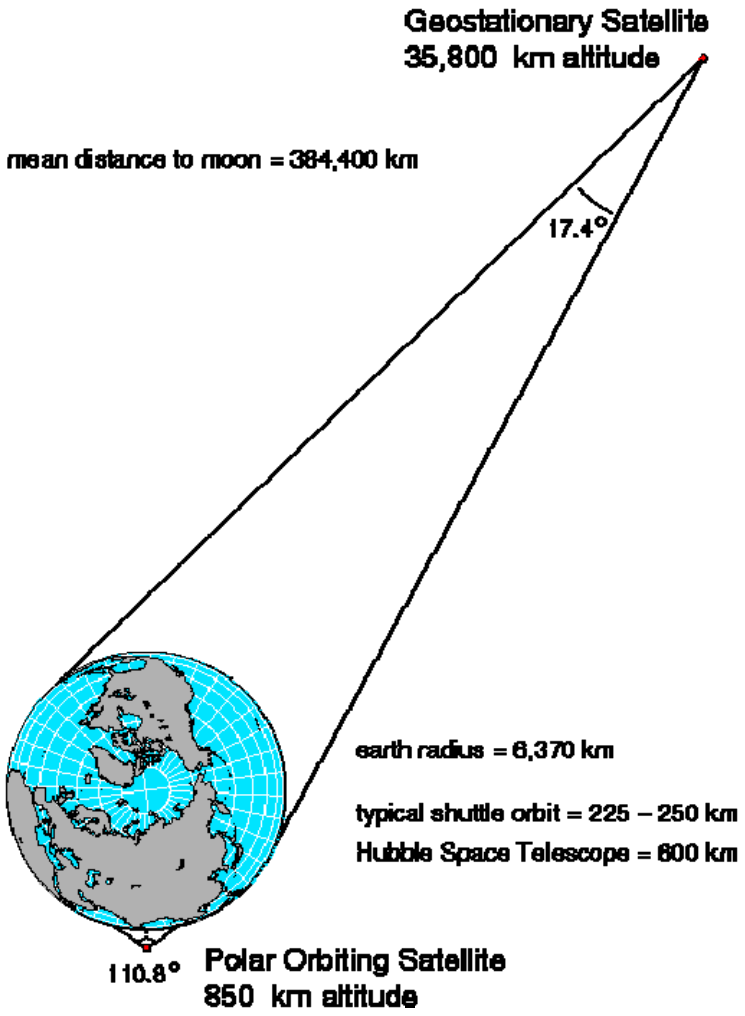


Figure 2.15 True relative distances from the Earth of geostationary and LEO satellites. (After Johnson, 2006.)

The most important meteorological geo-synchronous satellites are the GOES-East and GOES-West satellite families operated by NOAA, the Meteosat satellite family

operated by Eumetsat, the GMS satellite family operated by the Japanese NASDA, the INSAT satellite family operated by the Indian ISRO, and the Feng-Yun-2 satellite families operated by the Chinese Meteorological Agency.

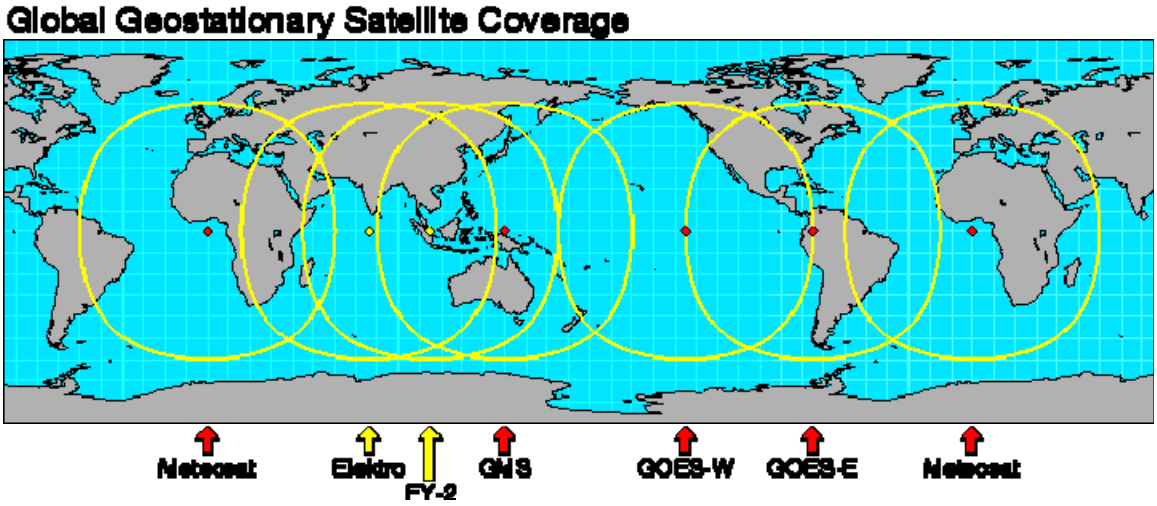


Figure 2.16 Coverage areas to six geostationary meteorological satellites currently in operation or in the planning stages. (After Johnson, 2006.)

The Meteosat First Generation Geostationary Satellites

Nominally located above the intersection between the Equator and the Greenwich Meridian (i.e. at 0° latitude and 0° longitude), the Meteosat satellites monitor the weather over the eastern Atlantic, Europe, the Middle East, Africa, and the western Indian Ocean (Figure 2.17).

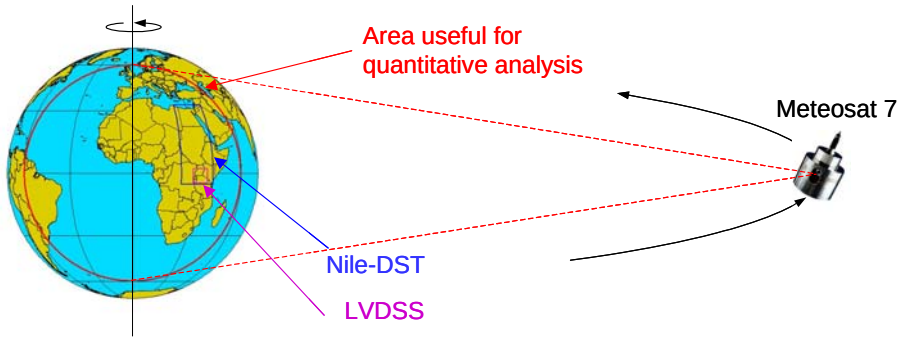


Figure 2.17 Area monitored by the Meteosat family of geostationary satellites.

The first satellite of the Meteosat series was launched on November 23, 1977 by a consortium of eight European countries, which later evolved into what is today the European Space Agency (ESA). The next satellite was launched in 1981 and, since then, Meteosat services have continued without major interruptions. In 1987 Eumetsat took over the management of the reception and distribution of the Meteosat data. In 2002, ESA launched the first satellite of the Meteosat Second Generation series, which carries a more powerful imager. Meteosat-7, the most recent satellite of the Meteosat First Generation series, will continue the 0° Longitude operational service until June 2006 (Eumesat, 2006).

The main instrumentation carried by the Meteosat First Generation satellites is the Meteosat Visible and InfraRed Imager (MVIRI), an imaging radiometer able to produce digital images of the full Earth disk in three bands of the electromagnetic spectrum, as described in Table 2.3.

IR and WV images are composed of 2,500 lines, each with 2,500 pixels, while VIS images have a higher spatial resolution (5,000 lines, each with 5,000 pixels). East-West scanning of each line is achieved through the rotation of the satellite (spin stabilization). The sequence of lines is produced by incrementally tilting the telescope from South to North. During each rotation, MVIRI scans a line composed by 2,500

pixels (two lines of 5,000 pixels each for the VIS channel). Scanning of the full Earth disk requires 25 minutes and is repeated every 30 minutes, beginning at the minutes 00 and 30. In the five-minute interval between the end of one scanning cycle and the beginning of the following one, the telescope is brought back to the initial position, recalibrated, and stabilized (Eumetsat, 2000).

Table 2.3 The Meteosat imaging radiometer channels. (Eumetsat, 2000.)

Spectral Band Name	VIS	WV	IR
Spectral band range	0.45 – 1.00 μm (visible)	5.70 – 7.10 μm (mid IR)	10.50 – 12.50 μm (thermal IR)
Instantaneous field of view at sub-satellite point	2.5 km	5.0 km	5.0 km
Lines per image	5,000	2,500	2,500
Pixels per line	5,000	2,500	2,500
Image duration	25 minutes	25 minutes	25 minutes
Image recurrence	30 minutes	30 minutes	30 minutes
Information provided	Daylight albedo	Water vapor amount in the middle troposphere (600 to 300 hPa)	Temperature of cloud tops and ocean surface
Application	Cloud, pollution, and haze detection; severe storm identification	Estimating regions of mid-level moisture content and advection; tracking mid-level atmospheric motion	Identifying cloud-drift winds, severe storms, and heavy rainfall; determination of sea surface temperature

Meteosat satellites carry an on-board blackbody mechanism to continuously monitor and correct the performances of the IR and WV channels. Further, comparisons of the observed radiances with surface-based radiation measurements from selected areas

of the Earth are carried out on a daily basis to improve instrument calibration (vicarious calibration).

Raw images are transmitted from the satellite to the Eumetsat Mission Control Center in Darmstadt, Germany, where they are corrected, rectified, and retransmitted back to the Meteosat satellites, which disseminate them to the user stations. The entire operation lasts 10-30 minutes. Images are available in digital format (High Resolution Imagery) or in analog format (WEather FAXsimile) images. Digital images are broadcast in two formats (Figure 2.18):

The A-format, which covers the full Earth disc and has full spatial resolution in the IR and WV channels, but only half spatial resolution in the VIS (2,500 lines each with 2,500 pixels).

The B-format, which gives priority to the European-Mediterranean sector, but provides VIS images at full resolution (5,000 lines each of 5,000 pixels).

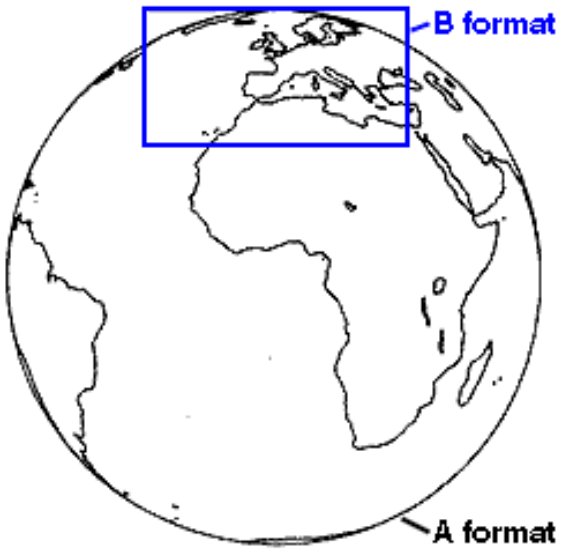


Figure 2.18 Meteosat High Resolution Imagery Formats. (After Eumetsat, 2000.)

Although images from the three bands are collected at full resolution for each time slot, it is possible to simultaneously disseminate only two of the three channels. The couple VIS-IR is transmitted during the daytime (between 5:30 GMT and 21:00 GMT), while the pair WV-IR is transmitted during nighttime (21:30 GMT to 5:00 GMT) and as an additional image in some of the daytime slots (Eumetsat, 1998). Further, transmission of the B-format is suspended during slot 23 (11:30 GMT).

2.3.2.2 Low-Earth Orbit Satellites

Low Earth Orbit satellites are placed at an altitude of 350-800 km in orbits inclined with respect to the equatorial plane (Figure 2.15). Orbits lower than this would rapidly decay because of the atmospheric drag, while the intense radiation and charge accumulation of the inner Van Allen belt would limit the life of satellites placed at higher orbits. The low satellite orbit affects image coverage in three ways:

First, LEO satellites can carry sensors operating in the microwave band because the short distance from the target allows obtaining a good spatial resolution even with antennas of limited dimensions.

Second, they cannot continuously monitor a given area. This is because LEO satellite velocity must be higher than the Earth's rotation speed to balance the strong gravitational attraction. With a typical velocity of around 8 km/s (27,400 km/h), LEO satellites circle the Earth in about 90-100 minutes.

The third effect on LEO images is that the portion of the Earth's surface that can be observed at any given time varies between tens and hundreds of kilometers in width. As a satellite revolves around the Earth, the sequence of patches observed at consecutives times forms a continuous band called a *swath* (Figure 2.19 A).

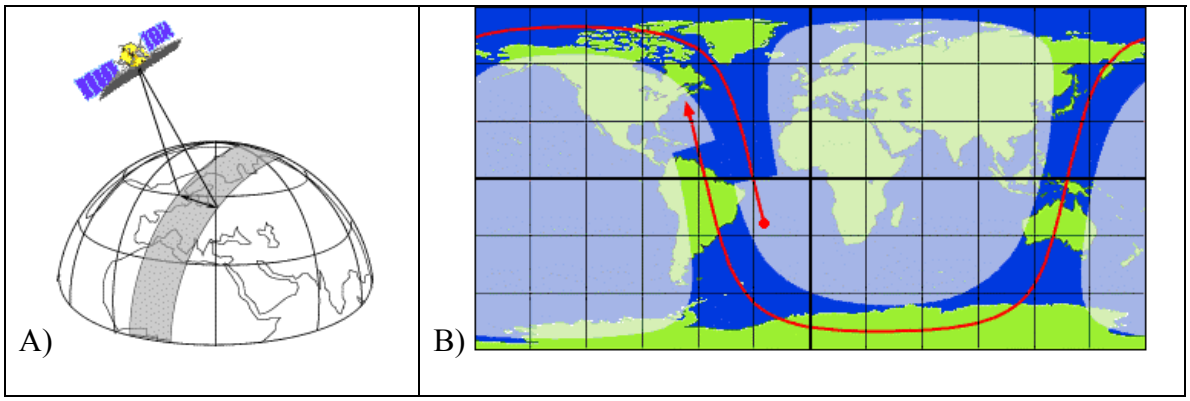


Figure 2.19 Characteristics of a LEO satellites orbit. A) The area observed during an overpass is called a "swath". (After Canada Centre for Remote Sensing, 2002.) B) Swath typically covered by a LEO polar satellite during an entire orbit. (After Eumetsat, 2002.)

After completing one orbit, LEO satellites do not pass over the same Earth location they were at the beginning of the orbit because in the 90-100 minutes it takes the satellite to make one orbit, the Earth has rotated eastward (Figure 2.19 B). This allows the satellite swath to cover new areas with each consecutive pass. A LEO satellite's orbit can be synchronized with the rotation of the Earth to monitor the entire Earth's surface in a given number of orbits. The interval of time required for the satellite to pass over a given geographical location is called a *revisit period*.

Although LEO satellite orbits may take a variety of inclination angles with respect to the equatorial plan, the one most commonly used is the "polar orbit", which nearly overpasses both poles. This orbit is so popular because it allows high frequency monitoring of the high latitude areas that are beyond the reach of the geostationary satellites. Further, if the polar orbit is tilted at the angle of 98.7° with respect to the equatorial plane, the angle between the orbital plane and the sun remains constant. This allows the satellite to cross a particular line of latitude always at the same local solar time and results in consistent lighting conditions (sun-synchronous orbits). This is due to the fact that at this inclination the orbit's precession matches the apparent motion of the sun as seen from Earth orbit, i.e. about one degree eastward each day.

Polar orbit satellites travel northward on one side of the Earth and then southward on the second half of its orbit. If the orbit is also sun-synchronous, the northward (or ascending) pass is most likely on the shadowed side of the Earth while the southward (or descending) pass is on the sunlit side. Sensors recording reflected solar energy can operate only during the descending pass, when solar illumination is available.

The Tropical Rainfall Measurement Mission Satellite

Launched on 27 November 1997, the Tropical Rainfall Measuring Mission (TRMM) satellite is a joint mission between NASA and the Japan Aerospace Exploration Agency (JAXA), aimed at providing a detailed and comprehensive dataset on the four dimensional distribution of rainfall, hydrometeor structure, and associated latent heating over under-sampled tropical oceanic and continental masses. The combination of precipitation radar, advanced PM sensors, and visible/infrared imager, the low-inclination orbit, and the unexpected longevity have allowed TRMM to meet and exceed all its research goals and to represent one of major successes of NASA's research program over the last decade (NASA, 2005). TRMM is scheduled to operate until 2009 (NASA-GSFC, 2006a).

TRMM is placed in an orbit inclined 35° with respect to the equator at an altitude of 350 km (402 since August 2001). This allows for a more frequent sampling (one to three times a day) of the precipitation in the tropics than that permitted by the sun-synchronous DMSP-SMM/I satellites (once a day for each satellite). Further, the time at which TRMM passes over a given location is not fixed (as for the DMSP-SMM/I satellites), but spans the entire twenty-four hours over a 42 day cycle. This characteristic is important because tropical precipitation undergoes a strong diurnal cycle that cannot be properly sampled by measures taken during only two periods of the day, as demonstrated by Li et al. (1996) and Morrissey and Janowiak (1996).

The main sensors dedicated to the study of precipitation are the TRMM Microwave Imager (TMI), the Visible and Infrared Radiometer System (VIRS), and the Precipitation Radar (PR) (Table 2.4 and Figure 2.20).

TMI is a nine-channel microwave radiometer with architecture similar to that of the Special Sensor Microwave/Imager (SSM/I). The main differences are the addition of two 10.7 GHz vertical and horizontal polarization channels, the shift of the frequency of the water vapor channel from 22.325 to 21.3 GHz, and the higher spatial resolution due to the lower orbit altitude. The antenna beam is inclined at a constant angle of 49°with respect to the nadir and observes an arc of 130° in front of the satellite. This results in a swath width of 758.5 km (Figure 2.20) with a spatial resolution varying from 63x37 km at 10.65 GHz to 7x5 km at 85.5 GHz (Kummerow et al., 1998). The TMI has operated perfectly since TRMM launch, with no drift or deterioration in its performances.

Table 2.4 Characteristics of TRMM sensors for precipitation study. (After NASA, 2005.)

Microwave radiometer (TMI)	Radar (PR)	Visible and infrared Radiometer (VIRS)
10.7, 19.3, 21.3, 37.0, and 85.5 GHz (dual-polarized except for 21.3: vertical only)	13.8 GHz	0.63, 1.61, 3.75, 10.8, and 12 μm
11 km X 8 km field of view at 37 GHz	5-km footprint and 250-m vertical resolution	2.5-km resolution
Conically scanning (530 inc.)	Cross-track scanning	Cross-track scanning
880-km swath	250-km swath	830-km swath

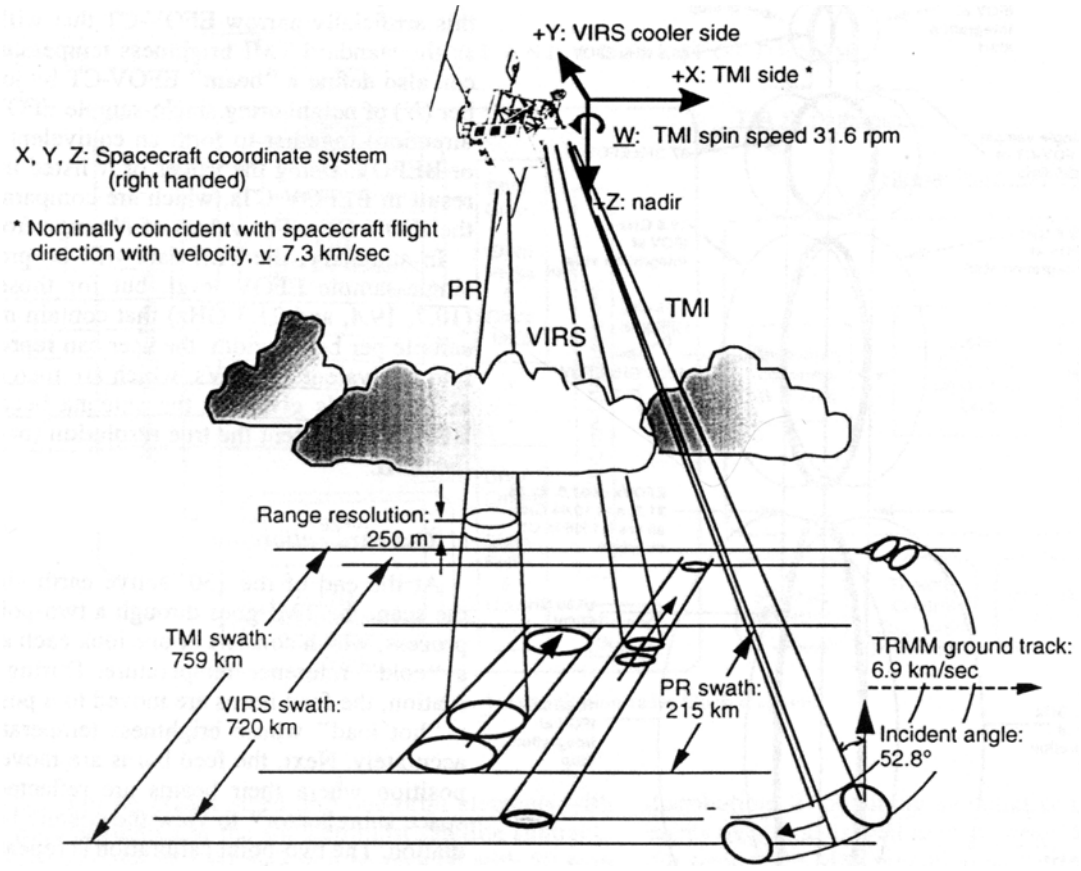


Figure 2.20 Schematic view of the scan geometries of the three primary rainfall sensors: TMI, PR, and VIRS. (After Kummerow et al., 1998.)

VISR is a five-channel imaging spectroradiometer based on the architecture of the Advanced Very High Resolution Radiometer (AVHRR) carried by the NOAA polar satellites. Due to its higher orbit elevation, VIRS has a lower resolution than AVHRR (2.11-km against 1.1-km at nadir), but the visible data of VIRS are more precise thanks to its onboard solar diffuser for post-launch calibration of the reflected solar bands (Kummerow et al., 1998). VIRS uncertainty for channel 1 and 2 reflectivity is around 6%, while the remaining three channels are more precise with an uncertainty limited to 3% (i.e., an uncertainty of ±2 K over a temperature of 300 K (Kummerow et al., 2000).

PR is the first rain radar carried by a satellite and will be the only rain radar in space until the next generation of GPM satellites, whose launches should begin in 2010.

Its main tasks are to detail the three-dimensional structure of rainfall, particularly the vertical distribution, and to obtain high quality, quantitative rainfall measurements over land and oceans. PR is a 2 m X 2 m, horizontally polarized, phased array antenna consisting of 128 slotted waveguides operating at the frequency of 13.8 GHz. The beam is electronically scanned cross track ±17° with respect to nadir (Figure 2.20). The swath is subdivided into 49 angle bins, with a near surface horizontal resolution of 4.3 km at nadir and a total width of 215 km. Each beam is subdivided into 80 range bins distributed every 250 m from just above the earth ellipsoid to an altitude of 20 km.

Short-term drifts in PR performances about once a day by an internal procedure, while long-term drifts are corrected with a ground-based Active Radar Calibrator (ARC) about four times a year (Kummerow et al., 2000). This process has guaranteed an absolute accuracy of less than ±0.5 dB and long-term relative stability of 0.1 dB (NASA, 2005).

PR algorithms for rain detection first determine whether each range bin is above of below the ground surface, if the corresponding data is present, and if it is reliable. After that, the presence of rain is assessed as "possible" (the return power is larger than the 90%-tile noise) or "certain" (the return power exceeds the sum of the "possible" noise level plus three times its standard deviation). PR classifies rain as stratiform, convective, or "other" by merging the results of two different methods (Product 2A23, NASA-GSFC, 2006b): one based on horizontal reflectivity distribution and called the *H-method*; the other based on the reflectivity vertical profile and called the *V-method*.

The H-method is a variation on the University of Washington convective/ stratiform separation method (Steiner et al., 1995). This method uses the horizontal pattern of the reflectivity Z at a given height (3 km at the tropics) to detect the presence of convective cores. The TRMM version, first classifies angle bins as part of a convective rain core if their maximum Z below the estimated 0°C level (Zmax) exceeds a given threshold or if Zmax stands out against the background reflectivity.

Table 2.5 Main TRMM precipitation products. (After NASA, 2005.)

Name	Reference No.	Purpose
Level2 data		
Surface cross section	2A21	Radar surface scattering cross section/total path attenuation.
PR rain type	2A23	Type of rain (convective/stratiform) and height of bright band.
TMI profiles	2A12	Surface rainfall and 3D structure of hydrometeors and heating over TMI swath.
PR profiles	2A25	Surface rainfall and 3D structure of hydrometeors over PR swath.
PR-TMI combined	2B31	Surface rainfall and 3D structure of hydrometeors derived from TMI and PR simultaneously.
Level-3 data		
TMI monthly rain	3A11	Monthly 5º rainfall maps-ocean only.
PR monthly average	3A25	Monthly 5º rainfall and structure statistics from PR.
PR-TMI monthly average	3B31	Monthly accumulation of 2B21 products and ratio of this product with accumulation of 2A12 in overlap region.
TRMM Multi-satellite	3B42	Multi-satellite precipitation data calibrated by TRMM, 3-hourly, 0.25º resolution.
TRMM Multi-satellite/et al.	3B43	3B-42 and et al. products-data merged into single rain product, monthly, 0.25º resolution.

Precipitation in convective cores and in the adjacent rain pixels is classified as convective. If the reflectivity is strong enough to be classified as certain, but not strong enough to be classified as convective, the rain type is set to stratiform. The "other" rain type is identified by pixels showing reflectivity below the "rain certain" level, including light rain and noise.

The V-method is based on the detection of the bright-band (BB) presence in those beams labeled as "rain certain". BB is present if the second derivative of the reflectivity Z with respect to range is negative (maximum in Z) and the Z above the BB peak decreases appreciably. Further, the height of the reflectivity peak must be located within 1500 m from the estimated freezing level, and it should be relatively constant in neighboring angle bins.

After assessing the existence of the BB, the V-method classifies the rain type according to the following criteria:

1. Rain is classified as stratiform whenever BB exists.
2. If the maximum value of Z for the angle bin exceeds a convective threshold, but BB is not detected, the rain type is classified as convective.
3. For all the cases when rain is present (i.e., reflectivity is strong enough), but 1. and 2. are not satisfied, the rain type is defined as "other".

The final classification of the beam in "convective", "stratiform", or "other" is obtained by merging the outcome of the V-Method and H-Method with other information on the presence of BB. The V-method output is emphasized in the classification of stratiform rain and the H-method output in the classification of convective rain (NASA-GSFC, 2006b).

Smearing of BB near the antenna scan edges seriously affects BB detection: the rate of BB detection is about 80% for antenna scan angles in the interval ±7° from nadir, but only about 20% at the swath edges. This problem is aggravated by the fact that the rain/no-rain discrimination also seems to have angle bin dependence (NASA-GSFC, 2006b).

At the PR frequency (13.8 GHz), the radar pulses are significantly attenuated by the precipitation encountered before reaching their target. TRMM Product 2A25 quantifies this effect for weak precipitation rates by adopting a variation of the Hirtschfeld–Bordan method for determining the Path Integrated Attenuation (PIA). For

high precipitation rates, the PIA is estimated by measuring the decrease in the power returned by the surface (apparent surface cross-section). PIA for intermediate precipitation is obtained by combining the two results (NASA-GSFC, 2006b; Meneghini et al., 2000; Iguchi and Meneghini, 1994).

2.3.3 Procedures for Remote Sensing of Precipitation

Although in the last fifty years researchers have developed hundreds of different algorithms for estimating precipitation using satellite images, procedures for estimating precipitation can be grouped into three broad categories:

1. Indirect schemes, using observation of cloud characteristics in the infrared (IR) and visible (VIS) spectrum
2. Direct schemes, utilizing images of the microwave radiation absorbed, scattered, and reflected by hydrometeors
3. Mixed schemes, combining both direct and indirect schemes

2.3.3.1 Indirect Schemes

Indirect schemes estimate precipitation based on information on the presence, typology, and evolution of the cloud systems. The basis of these techniques is that the presence of clouds is a necessary condition for precipitation to occur in appreciable amounts. Further, as pointed out in section 2.1, clouds of different types tend to generate precipitation in specific ranges of intensity. For example, tall, but thin, cirri do not generate precipitation, while tall and thick cumulonimbi are frequently associated with heavy convective precipitation (Section 2.1.1).

The large majority of indirect schemes use satellite images in the thermal infrared band (10-12 μm, IR) that, as discussed in Section 2.3.1, can be related to the presence and

elevation of clouds. In the tropics and during the summer months also at mid-latitudes, cold clouds are frequently associated with convective clouds that produce heavy rains (Section 2.1.1). However, even in these favorable situations, the relation between low IR temperature and precipitation is far from being exclusive, since some cold clouds do not produce rain (for example, cirri and clouds in the leeward side of mountain ranges), and some warm clouds can produce significant amounts of rain (for example, shallow stratocumuli and clouds caused by orographic and coastal lifting).

Table 2.6 Summary of indirect satellite estimation methods.

Method	Sensors	Temporal resolution		
		< 1 Day	1 – 10 Days	Monthly
IR Thresholding	IR		TAMSAT CCB4	GPI
Life history	IR	Griffith-Woodley IFFA Vincente et al. CST		
Bi-spectral	IR and VIS	Tsonis et al. RAINSAT		
Multispectral	IR+2 or more channels	Kurino	Georgakakos et al.	

Stronger than the relation between low IR temperature and precipitation is the relation between precipitation and cloud's albedo (King et al., 1995), since brighter clouds are also thicker, and thicker clouds are more likely to generate intense rain (Section 2.3.1). However, the absence of VIS data during nighttime means that albedo is generally used only to complement the information provided by IR data during the daytime period (King et al., 1995; Tsonis et al., 1996).

The various indirect schemes that have been used over the years differ mainly in the way they translate IR/VIS data into precipitation amounts and in the way they use auxiliary data to improve cloud type identification.

Traditional indirect techniques can be usefully categorized according to the scheme shown in Table 2.6.

IR-Thresholding Methods

The simplest approach to estimating precipitation with indirect methods considers as rainy only the pixels with IR temperature below a certain threshold and assigns them a given rain rate. The most well-known of these methods, and the most popular precipitation estimation technique, is the GOES Precipitation Index, or GPI, (Arkin and Meissner, 1987). This and similar methods use frequent IR images from geostationary satellites to determine the mean portion of a 2.5ºx2.5º area covered by clouds with IR temperature below 235ºK (cold clouds). The mean monthly precipitation over the area is then obtained by multiplying this fractional coverage by the fixed precipitation rate of 3.0 mm h^{-1}. That is:

$$GPI = GFcT = G\iiint i_c(x, y, t, Th_{IR})dxdydt \qquad (2.3)$$

where: GPI = mean area precipitation for period T in mm

G = empirical rain rate, set to 3 mm h^{-1}

Fc = mean fractional coverage of clouds colder than Th_{IR} during period T

T = length of the time period (hours)

$i_c(x,y,t,Th_{IR})$ = 1 if $IR(x,y,t) < Th_{IR}$, 0 otherwise

The basis of this technique is the pioneering works of Arkin (1979) and Richards and Arkin (1981) who found a strong linear relation between fractional coverage of cold clouds and mean areal precipitation over the eastern Atlantic. The relation was consistent

only for areas larger than 0.5ºx0.5º and improved with the extension of the area and period considered. They also found that the relation was relatively insensitive to the IR threshold in the range 225-255ºK that was used to define the cold clouds, a fact later confirmed by Menz (1997) for Kenya.

GPI and similar methods are popular because they are simple to implement, require readily available IR images, and have been demonstrated effective in estimating precipitation in regions where convective systems are dominant, such as the tropics. In these conditions, the cold-cloud fractional coverage of a region is strongly related to the Area-Time-Integral (ATI) of the area covered by precipitation, as pointed out by Atlas and Bell (1992). The ATI single storms are, in turn, directly related to the volume of rain produced by the storms (Doneaud et al., 1981; Lopez el al., 1983). That is:

$$\iiint r(x,y,t)dxdydt = R\iiint i_r(x,y,t)dxdydt = R * ATI \tag{2.4}$$

where: r(x,y,t) = rain intensity for pixel (x,y) at time t

i_r(x,y,t) = 1 if r(x,y,t)>0, 0 otherwise

R = conditional climatological rain rate

The combination of these two factors explains the success of such a simple method in estimating precipitation over large areas and long periods. Atlas and Bell (1992) also showed that the constant precipitation rate used in the GPI is proportional to the conditional climatological rain rate:

$$G = R\frac{ATI}{F_c T} = R\frac{\iiint i_r(x,y,t)dxdydt}{\iiint i_c(x,y,t,Th_{IR})dxdydt} \tag{2.5}$$

where: F_c = Mean cold cloud fraction

T = Estimation duration (hr)

$I_c(x,y,t) = 1$ if $IR(x,y,t) < Th_{IR}$, 0 otherwise

Over large space and time scales, the sample R approximates the true conditional climatological rain rate, while the ratio between cold cloud fraction and ATI is representative of the typical storm structure. Reducing the temporal and spatial scale, however, deteriorates both the relation between the IR<235º fraction and the ATI, and the probability that storms are completely confined within the area/period of interest, thus decreasing the method's reliability. Further, GPI does not perform well when the relation between cold-cloud area and ATI is not strong, such as for frontal and orographic rain (Arkin and Meissner, 1987; Adler et al., 1993; Thorne et al., 2001).

One of the major problems of GPI is that even where the correlation between cold-cloud coverage and precipitation is strong, the optimal IR threshold and rain rates that characterize the method vary seasonally and spatially (Arkin and Meissner, 1987; Johnson et al., 1994; Menz, 1997), requiring local calibration for a quantitative use of the technique.

While the cloud-top temperature provides sufficient information to estimate precipitation at coarse spatial and temporal resolution, it is insufficient at fine temporal and spatial scales, even in areas where convective precipitation is dominant (D'Souza et al., 1990). Several authors have used spatially and temporally varying thresholds to adapt the GPI concept to finer temporal and spatial scales. In the Tropical Applications of Meteorological Satellites (TAMSAT), the daily or dekadal precipitation at the pixel level is set as zero if the IR is always above a local optimal threshold. Otherwise, it is estimated as a linear function of the number of hours during which clouds colder than the threshold are present. The three optimal parameters of this relation are determined by calibration of the satellite data against rain gage measurements within climatically similar zones (Thorne et al., 2001). As threshold, the TAMSAT method selects the value that maximizes the probability of correctly identifying wet and dry periods and that balances

the probabilities of incorrectly classifying wet and dry periods. TAMSAT has been operationally used for more than twenty years in tropical and sub-tropical Africa for drought warning and crop prediction.

Similarly, the "B4:" technique (Bellerby and Barrett, 1993; Todd et al., 1995, 1999) defines a pixel as rainy if its IR is below a locally and temporally varying threshold. The daily precipitation over each rainy pixel is computed using a local conditional climatological daily rain, the pixel's cold cloud duration, and a weighting factor accounting for the differences between gage data and satellite estimates. Both the IR thresholds and the weighting of the cold cloud duration are continuously calibrated using the previous ten-day period for tracking seasonal variations.

IR-Life History Methods

Another technique used to improve precipitation estimation based solely on IR images, is to derive additional information on the cloud's dynamics by manipulating them. Many of these techniques are based on the observation that in convective storms, the precipitation intensity varies with the stage of the storm's *life-history*, with intense convective rain dominating the mature and developing phases, whereas weak stratiform rain is prevalent during the dissipating phase (Figure 2.7). Scofield and Oliver (1977) implemented this concept with a simple decision-tree technique based on the following conditions:

1. Bright clouds have a higher probability of rain
2. Intense precipitation is more likely when clouds show cold top temperatures and are in growing/cooling phases. Shrinking and warming clouds have a higher probability of producing light rain.
3. Rainfall is concentrated in the leading edge of a cloud system.

Given its good results, the approach was adopted as the basis for the Interactive Flash Flood Analyzer (IFFA), which is used operationally by NOAA to estimate precipitation from convective systems (Scofield, 1987). This procedure analyzes sequences of images from the geostationary satellite GOES and provides an estimate of precipitation every three hours. However, the applicability of this procedure is limited to one convective system at a time due to the heavy involvement of meteorologists in identifying the convective clusters and comparing their development with previous conditions.

Several authors have formulated automatic procedures for estimating precipitation based on the life-history concept. In the Griffith-Woodley technique and its variations (Griffith et al., 1976; Griffith, 1987), the precipitation volume from individual raining clouds (identified as pixels with IR<253 ºK) is estimated as the product of the cloud area, a factor inversely proportional to the cloud top temperature, and a function of the ratio between the cloud area and the maximum area experienced by the cloud system during its lifetime. The latter ratio is weighted differently if the cloud area is increasing (developing and mature phases) or decreasing (dissipating phase).

NOAA has developed an automated version of IFFA called the Auto Estimator (Vincente et al., 1998; Ferraro et al., 1999). This algorithm estimates the precipitation rate as a function of the thermal IR temperature, available moisture, and the cloud's growth rate:

$$
\begin{aligned}
&P' = 1.1183 * 10^{11} * e^{(-3.6383 * 10^{-2} * T^{1.2})} \\
&GR = 0 \text{ if } Tcold(t) > Tcold(t-1), 1 \text{ otherwise} \\
&PWRH = PW * RH, \text{empirically rescaled so that } 0 <= PWRH <= 2 \\
&P = \begin{cases} 72.0 * GR \text{ if } T < 200\ K \\ P * GR \text{ if } 200 < T < 210, PWRH > 1.0 \\ P * GR * PWRH \text{ otherwise} \end{cases}
\end{aligned}
\tag{2.6}
$$

Where T = 10.7 μm brightness temperature (k)

PW = Precipitatible water between the surface and the 500 mb (in)

RH = Mean relative humidity between the surface and the 500 mb (fraction)

Tcold(t) = IR temperature of the coldest IR pixels in the cloud system at time t (K)

The philosophy of this algorithm is that all precipitation is higher if the environment is moist and lower when it is dry. The dryness of the lower troposphere can be assessed from precipitable water and relative humidity data collected by rawinsondes. Further, precipitation occurs exclusively during the growing and mature stages of the storm. A storm is considered to be in growing or mature stages if the coldest pixels in the storm are as cold or colder than they were in the previous time step. Depending on how these pixels are defined, however, the algorithm could ignore the large areas of stratiform rain that are in the wake of mesoscale convective systems, which constitute a sizable portion of the total rain (Section 2.1.1). Despite this, the algorithm tends to overestimate the daily precipitation rate and the precipitation from slow-moving cold-topped mesoscale convective systems (Vincente et al., 1998).

A technique that explicitly separates convective rain from stratiform rain, indirectly accounting for a storm's life history is the Convective Stratiform Technique (Adler and Negri, 1988). This procedure considers as potential convective cores the spatial minima in the thermal IR (10.5-12.6 μm) images with temperature below 253 °K. Cold, but non raining, cirri are empirically screened out based on their local IR slope: a local IR minimum is considered a cirrus if the local temperature slope is shallow. That is, if:

$$T_{avg}-T_{min}>0.568(T_{min}-217.0) \quad (2.7)$$

Where: T_{avg} = average IR at the four pixels adjacent to the IR minimum location

Based on the one-dimensional cloud model of Adler and Mack (1984), convective precipitation of intensity R is assigned to an area Ar surrounding the IR minima which pass the cirrus screening test:

$$\begin{cases} T_C = 0.717 T_{\min} + 56.6 \\ R = 74.89 - 0.266 T_C \\ A_r = \exp(15.27 - 0.0465 T_C) \end{cases} \quad (2.8)$$

For each convective core, the temperature T_B of the surrounding stratiform anvil is computed as the modal IR temperature of the pixels with IR<253°K contained in an 80x80 km square centered on the convective core. The weighted average of the T_B temperatures of convective cores present in the region, TS, is used to identify the stratiform pixels. Pixels colder than TS and not belonging to the convective cores as defined in equation 2.8 are assigned a precipitation rate of 2 mm h^{-1}.

Bispectral Methods

Precipitation rates are more strongly related to visible radiation (VIS) than to the thermal infrared since brighter clouds are also thicker, and thicker clouds are more likely to generate intense rain (King et al., 1995). However, as previously stated, the absence of VIS data during nighttime means that albedo is generally used to complement the information provided by IR data during the daytime period.

Tsonis and Isaac (1985) devised a method for tracing instantaneous rain areas from VIS and IR images. Their technique is based on their observation that peaks in the bivariate frequency distribution of contemporaneous VIS/IR images correspond to different classes. Peaks associated with raining clouds tend to cluster in a well-defined low-IR/high-VIS region of the VIS/IR domain, which makes them distinctive from other

classes. After a peak (VIS_P, IR_P) has been identified as corresponding to rainy clouds, all the pixels having VIS larger or equal to VIS_P are classified as rainy. Average precipitation rate is then determined as a linear function of the rainy area, VIS_P, IR_P, $f(VIS_P, IR_P)$, mean VIS in the rainy area, and narrowness of the peak (Tsonis et al., 1996).

A simpler approach was taken by Lovejoy and Austin (1979). By using contemporaneous radar and satellite images, they created a look-up table for instantaneous precipitation probability as a function of IR and VIS values. This approach was later expanded in the RAINSAT algorithm, which included a look-up table for precipitation rates as a function of IR and VIS and data-smoothing spline filtering (King et al. 1995). In this algorithm the daytime single pixel instantaneous precipitation was obtained from the IR and VIS radiation, while the nighttime precipitation was a function of the IR alone. Despite its simplicity, RAINSATwas the best performing retrieval method in the AIP/1 intercomparison project (Arkin and Xie, 1994) and was among the best in the subsequent AIP/2, and AIP/3 (Ebert et al., 1996). The VIS data proved particularly effective when precipitation was from warm, orographically induced, clouds (King et al., 1995).

Multispectral Methods

Although most indirect schemes use thermal infrared and visible images, several precipitation estimation algorithms have used images in other bands, especially, but not exclusively, in conjunction with data from the polar orbiting satellites of the NOAA series, which have carried five-channel imagers since 1981 (NOAA-7).

Inoue (1987) found that cirrus and cumulus clouds could be identified exploiting the differences in infrared emissivity of ice and water particles. Inoue's method used the difference between the cloud's radiation at 11 and 12 μm to differentiate the two phases.

Turpeinen et al. (1987) incorporated data from the 6.3 μm channel of the Meteosat satellite (Water Vapor) to improve a GPI-like algorithm. These data are related to the upper-tropospheric humidity (Section 2.3.2.1) and have a role similar to that of the PWRH factor in equation 2.6.

Kurino (1997) took advantage of these relations to estimate instantaneous precipitation rates using images from the geostationary satellite GMS-5, which carries a four-channel imager. Kurinos' method is similar to that of the RAINSAT algorithm in that the probability of rain and the precipitation intensity are determined through a look-up table created by matching satellite images and contemporaneous radar data. The indexes of these look-up tables are the infrared brightness temperature at 11 μm (TB_{11}), the IR TB difference between 11 and 12 μm (TB_{11-12}), and the IR TB difference between 11 and 6.7 μm ($TB_{11-6.7}$). TB_{11} is, of course, tied to the presence and top-elevation of clouds. On the other hand, $TB_{11-6.}$ can also be used to identify deep convective clouds: TB_{11} is normally warmer than $TB_{6.7}$, but during deep convection the two temperatures are very similar. Further, convective cells upshots can sometimes protrude into the tropopause, making TB_{11} lower than $TB_{6.7}$. The author claims that this approach performs far better than using GPI or CST. A potential advantage of the method is that the 12 and 6.7 μm data are available for the entire day.

Georgakakos et al. (2000) used the bivariate IR/VIS frequency peak approach of Tsonis and Isaac (1985) to determine daytime rainy pixels in the Nile River watershed. During nighttime, they used the same approach, but applied to the Meteosat IR/WV domain. Estimate of daily rainfall was based on dynamic regression relationships between gage data and the number of wet half-hours determined in the first step.

2.3.3.2 Direct Schemes

Section 2.3.1 illustrates the theoretical basis for the estimation of precipitation in the microwave region (20 – 100 GHz). Here, it is sufficient to recall that non-

precipitating clouds are virtually transparent at frequencies below 40 GHz, while microwave radiation is strongly related to the presence of water particles trough absorption, emission, and scattering. Below 22 GHz, absorption is the dominating effect, while scattering dominates the region above 60 GHz. Either process can be present at intermediate frequencies (Kidd and Barrett, 1990). Precipitation retrieval over the ocean is based on the radiation emitted by the brighter rain droplet clouds. Given the uniform background and the strong emission difference between raindrops and the ocean, even simple thresholding techniques prove effective in identifying precipitation. Over land, precipitation retrieval based on emission is more difficult because of the higher and inhomogeneous radiation from the land surfaces. Consequently, early algorithms, based on thresholding of single-frequency radiation, were effective only for the identification of heavy rain (Kidd and Barrett, 1990). More recently, high frequency (85 GHz) sensors have allowed the estimation of precipitation by exploiting the scattering properties of raindrops and ice particles, which decrease the upwelling radiation emitted by the Earth's surface (Kidd, 2001).

Although other classification schemes are possible, modern retrieval can be classified into two major groups (Anagnostou, 2004). The first group is composed of the "physically-based" retrieval techniques such as those by Olson (1989), Smith et al. (1994), and Kummerow el al. (1996). These algorithms generate in advance a large database of hydrometeor profiles of the atmosphere and associated brightness using radiative models. During the estimation phase, they select and adapt the profiles that most closely match the observed radiation. These algorithms are used mainly over the oceans, where the cold and uniform background does not mask the lower frequency channels, yielding a unique solution to the inversion problem. The second group of retrieval methods is composed of "statistically-based" algorithms. These algorithms derive their radiation-precipitation relationships from comparisons of satellite radiation with collocated precipitation measurements at the surface (e.g., Olson et al., 1991;

Ferraro et al., 1994; Kidd, 1998) or from other satellite sensors (e.g., Grecu and Anagnostou, 2001; McCollum and Ferraro, 2003). Results from most of these algorithms are very similar, because of the strong relation with the precipitating particles (Kidd, 2001), and much better than the equivalent estimates of instantaneous precipitation from indirect methods (Ebert et al. 1996).

The problem with the direct estimation schemes is that up to now, microwave sensors are carried only aboard LEO satellites, making the continuous monitoring of a region impossible in large areas of the world. Polar orbiting satellites, such as the DSMP satellites carrying the SSM/I microwave sensor, have a revisit frequency at the equator of less than twice a day. Thus, a fleet of two satellites would be able to provide precipitation estimates less than four times a day. Actually, Smith (2001) estimated that at the time, the entire fleet satellites carrying PM sensors had a maximum revisit period for the tropical belt exceeding nine hours. Kidd et al. (2003) reported that, in certain situations, SSM/I may pass over a region just once a day.

Despite the poor sampling frequency, the estimation of monthly precipitation rate over large areas would still be relatively accurate. Bell et al., (1990) estimated that TRMM PR radar, which has a poorer revisit period than TMI or SSM/I, could measure monthly precipitation over 500x500 km areas with a sampling-related uncertainty of less than 10%. Such an estimation was obtained using a stochastic model of precipitation based on the GATE data. On the other hand, Steiner and Houze (1998), using actual radar data, showed that the TRMM PR radar estimates of monthly precipitation over 500x500 km areas had a more realistic sampling-related uncertainty of about 20%. In the case of measurements taken from polar-orbiting sun-synchronous satellites, these errors may be underestimated, because these satellites pass over the same latitude always around the same time. At low sampling frequency, this results in a mischaracterization of the strong diurnal cycle of tropical precipitation (Li et al., 1996; Morrissey and Janowiak, 1996).

Estimation of daily or dekadal precipitation is affected so much by the low temporal sampling frequency, that the GPCP AIP intercomparison did not even consider direct schemes for this type of precipitation (Ebert, 1996).

The sampling deficiency of direct schemes should be corrected in the period 2010-2012 with the launch of a new series of satellites carrying passive and active microwave sensors established by the Global Precipitation Measurement protocol. These satellites should able to guarantee a revisit period of just three hours.

2.3.3.3 Mixed Schemes

Indirect schemes have the advantage that they can exploit the almost continuous flow of images provided by the geostationary satellites. However, the relation between IR/VIS images and precipitation is indirect and incomplete. In order to overcome these drawbacks, a continuous recalibration in space and time is necessary to improve model performances, requiring access to good networks of rain gages/radars.

Direct schemes are much better at measuring instantaneous precipitation rates, but this benefit is hindered by the poor temporal frequency of microwave data, which, in the tropics, is presently below one image every six hours.

Mixed schemes try to overcome these deficiencies by combining the two types of procedures. In the majority of mixed schemes, coincident IR/VIS and PM precipitation images are matched and used to update/localize the IR/VIS-precipitation relationship. The updated relationship is then applied to the full flow of the IR/VIS images to produce high temporal resolution estimates.

Some of the combined schemes use microwave data only to continuously recalibrate the parameters of traditional indirect schemes. This is the case of the Adjusted GPI (Adler et al., 1993, 2000), Universally Adjusted GPI (Xu et al., 1999), Threshold-Matched Precipitation Index (Huffman et al., 2001), and CST/TMI (Negri et al., 2002). The objective of the first algorithm is to use the average precipitation rate

observed by TRMM in place of the fixed 3.0 mm h^{-1} in the GPI algorithm. AGPI obtains this goal by computing the following adjustment coefficient:

$$r_a = \frac{V_M}{V_{MIR}} \tag{2.9}$$

Where: V_M = total monthly rainfall measured by the TRMM Combined Instrument (TCI) algorithm for a given area

V_{MIR} = total monthly rainfall estimate by GPI using only IR data coincident with the TRMM data in the area

The adjustment coefficient r_a is then used to correct the mean average precipitation estimates produced by GPI during the month. The adjustment is done at a 1° x 1°spatial resolution.

The Universally Adjusted GPI algorithm takes this approach a step further, by varying in time and space both the GPI rain rate and the optimal threshold so that the total error between the monthly IR-based and TRMM-observed precipitation is minimized. The procedure for each target area can be outlined as follows (Xu et al., 1999):

1. Calculate the number of rainy pixels R_M and the mean rate p_M from the microwave data;
2. Determine the optimal IR temperature threshold T^* that minimizes the absolute difference between the number of rainy pixels R_M and the number of collocated pixels colder than T^*, $N_M(T^*)$;
3. Compute the cold cloud fractional coverage $F_C(T^*)$ for each hour using all available IR data for that time-step;
4. Multiply $F_C(T^*)$ by p_M to obtain an estimate of the hourly MAP over the area;

5. Cumulate hourly precipitation as necessary to obtain the precipitation for the desired period.

Xu et al. (1999) show that by adjusting the IR threshold, UAGPI can tackle cases dominated by "warm" rain, for which no-IR temperature is below 235 °K. Further, because of the higher flexibility allowed by using two adjustment parameters, UAGPI performs consistently better than AGPI. TMPI is a version of the UAGPI algorithm, featuring a conceptually similar, but more complex, statistic algorithm for determining the optimal thresholds and precipitation rates (Huffman et al., 2001). The UAGPI is used by the Global Precipitation Climatology Project to provide daily and 3-hr precipitation rates at 1°x1° resolution.

In the CST/TMI, the Cumulus/Cirrus screen that was empirically defined as equation 2.7 in CST is recomputed using TRMM TMI data. TMI-retrieved rain rates and rain types were used to classify each local IR minima of the training set as either convective or non-convective. The discriminating lines in the (T_{min}, T_{avg}- T_{min}) domain separating cumulus clouds from cirrus clouds were determined by minimizing the total entropy of two groups of T_{min}. The total area of convective rain and the mean convective rain rate from TMI images were used to compute the correspondent parameters in equation 2.8. The total area of stratiform rain and the mean stratiform rain rate were instead used to calibrate the IR threshold separating stratiform pixels from no-rain pixels and the precipitation rate to be assigned to the former ones.

Other mixed schemes have been explicitly designed to exploit the combination of IR/VIS and PM data. The MIRA scheme (Todd et al., 2001; Kidd et al., 2003) uses a Probability Matching Method technique to produce precipitation as a function of the IR temperature. In PMM the cumulative histogram of observed precipitation rates is matched with the inverse cumulative histogram of coincident IR temperatures. This means that the highest 10% of precipitation rates will be associated to the coldest 10% of

IR pixels, the highest 20% of precipitation rates to the coldest 20% of IR, etc. In Todd et al. (2001) the matching was done over areas of 1°x1° and periods of one month. In Kidd et al. (2003), the matching is done over five-day periods and 1°x1°areas, although a 5°x5° Gaussian filter is used to smooth the histograms. A fundamental assumption of this approach is that precipitation rates are inversely related to IR temperature. While this assumption is relatively sound in environments where convection is the exclusive lifting mechanism, it is erroneous in regions where "warm" rains are common, like mountainous areas and seashores. In these cases, the PMM approach tends to overestimate the precipitation from cold clouds and underestimate that from warm clouds.

Hsu et al. (1997, 1999) and Bellerby et al. (2000) have proposed two similar more sophisticated approaches to the joint use of TRMM measurements and geo-stationary IR/VIS data. In both approaches, hourly rainfall rates over small areas (in the order of 0.12° x 0.12°) are estimated from contemporaneous IR/VIS geo-stationary images using specifically trained neural networks. In Hsu et al. (1999) the output of the neural network scheme is the 30-minute rainfall rate for a square of 25 pixels covering 0.25° x 0.25°. The inputs of the neural network is the IR spatial texture, which is represented by the IR temperature in the central pixel of the square, the mean and variance of IR temperature over the central 3x3 pixel square, and the mean and variance of IR temperature over the whole 5x5 pixel square. During daytime the VIS texture adds to the input similar statistics for the VIS channel. A neural network-based scheme (called PERSIANN) first classifies this set of statistics into a large number of groups associated with different cloud characteristics. Then, a multiple linear regression function specific to each group relates the input set to the 30-minute rain rate. Finally, the neural network output is aggregated over 1° x 1° areas and daily intervals. Whenever a new measure of the difference between the TMI 2A21 rainfall estimate and the PERSIANN estimate is available, the coefficients of the multiple linear regression functions are updated. PERSIANN is considered one of the leading mixed schemes.

Bellerby et al. (2000) extended this approach by adding the texture of GOES channels 2 (3.9 μm), and 3 (6.7 μm) to the input set used in PERSIANN. Further, they recognized that the temporal evolution of cloud systems is also important for assessing

rainfall and used the statistics of two consecutive GOES images as input of the Neural Network. Additional parameters were also considered for a total of 45 input variables. A two-layer feed-forward artificial neural network, composed of 200 and 100 nodes, was trained against coincident TRMM PR data to produce the 30-minute 0.12° x 0.12° rainfall. This approach, however, must not have been particularly successful, since it has not been presented again in other papers.

All mixed schemes presented previously, use coincident IR/VIS and PM data to update the IR/VIS-precipitation relation to be used in an indirect scheme. A completely different approach to merging direct and indirect methods is taken by Joyce et al. (2004). Their Climate Prediction Center morphing (CMORPH) method instead uses motion vectors derived from half-hourly IR imagery to propagate in time and space the precipitation rates observed by PM sensors. The displacement vectors governing the rain propagation are determined as those maximizing the correlation between the IR image at time t and a spatially shifted version of the IR image at t+1. PM images are propagated both forward and backward in time using the displacement vectors. The shape and intensity of the precipitation features for the interval between consecutive PM images is then obtained as a time-weighted linear interpolation of the forward-propagated older image and the backward-propagated newer image (morphing).

A caveat of this method is that PM images must be available at a frequency of at least one image every six hours. Below such a frequency, the performances decrease dramatically. CMORPH is able to obtain good results because it integrates the PM images from TMI and SSM/I with those produced by the AMSU-B sensors aboard the NOAA-15, NOAA-16, and NOAA-17 satellites, thus guaranteeing revisit periods generally below three hours.

2.3.3.4 Performances of Remote Sensing of Precipitation

Despite all the advances in available sensors and methodologies, remote sensing of precipitation is still affected by large errors. Further, performances of estimation algorithms vary a lot according to where they are applied.

Although relatively old, the results of the AIP comparison give a global picture of the variability in estimation performances at a variety of scales (Ebert et al., 1996).

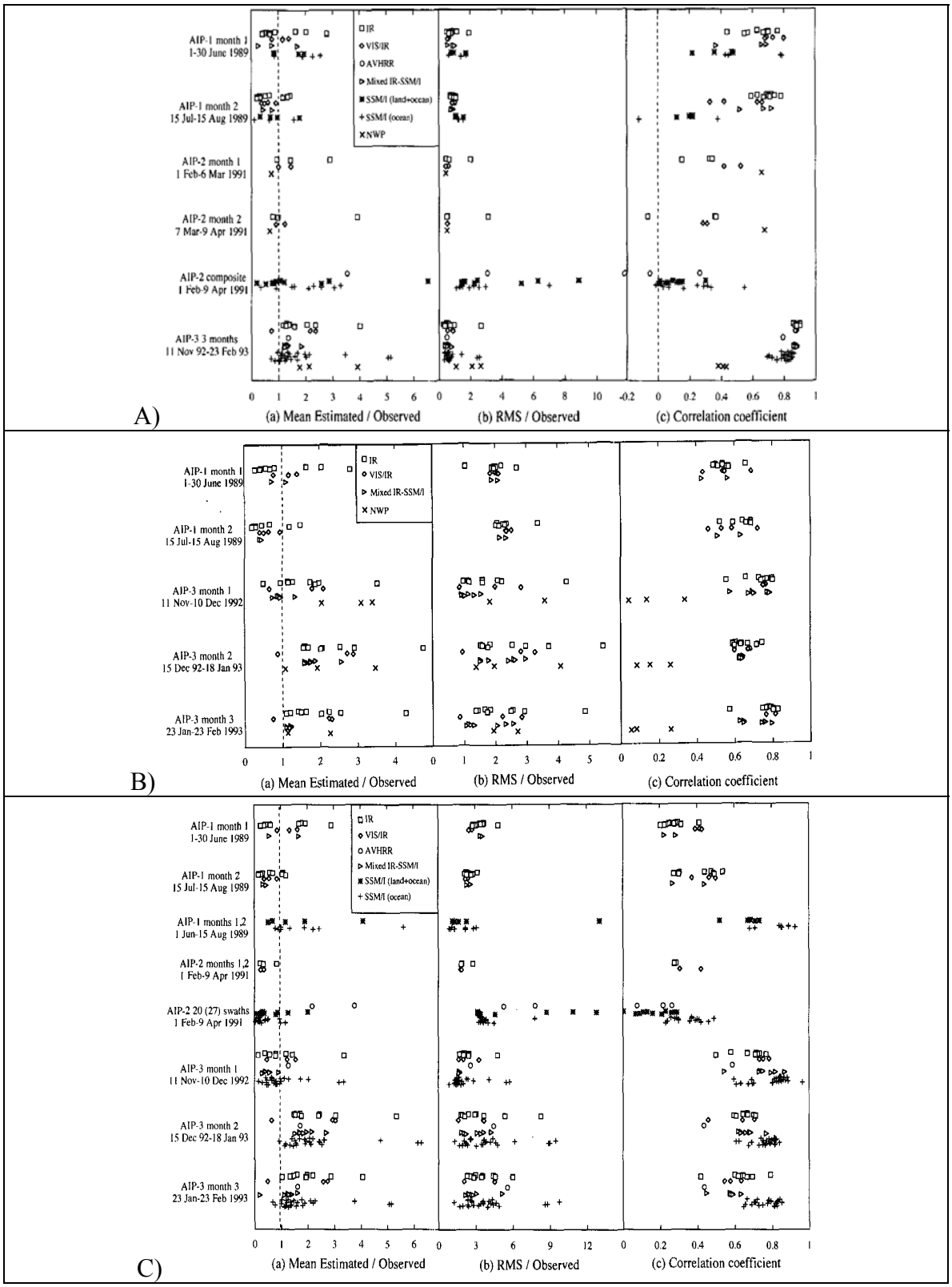


Figure 2.21 AIP validation statistics for A) Monthly rainfall; B) Daily raifall; C) Instantaneous rainfall. (After Ebert et al., 1996.)

The three AIP experiments took place over Japan (AIP/1), western Europe (AIP/2), and the TOGA COARE area in the western Pacific Ocean (AIP/3). Figure 2.21 reports algorithm performances in estimating mean area precipitation at 1.25°x1.25° resolution (AIP/1 and AIP/2) or 0.5°x0.5°resolution (AIP/3) and different time resolutions. Chart A) shows that monthly precipitation estimates have correlations varying from above 0.8 in strongly convective environments to 0.2-0.6 in mid-latitude spring conditions (AIP/2). Bias may easily be above 100%, while RMSE is generally less than the mean precipitation rates, although higher figures are not uncommon. Chart B) points out that the RMSE/mean daily precipitation is generally above 100%, while correlation can reach 0.8 in optimal conditions (tropical ocean), but declines rapidly in more complex environments. Chart C) confirms the superiority of PM schemes in determining instantaneous precipitation rates.

More recent algorithms perform generally better, but even they are far from perfect. Adler et al. (2001) investigate the ability of remote sensing algorithms in evaluating monthly precipitation at 2.5°x2.5° resolution. The satellite estimates were compared with gage data from dozens of test sites around the world.

Figure 2.22 shows that the satellite-gage correlation is generally between 0.7 and 0.8 over tropical oceans and tropical land, but below 0.6 over continental mid-latitudes. Only the gpm and nmg algorithms, which are continuously calibrated against rain gage data, perform better over land. The bias/mean precipitation is generally below 50%, but values higher than 25% are common. Note that the RMSE values over continental mid-latitudes are lower than RMSE over the tropics because mid-latitude precipitation is also lower. A chart reporting the RMSE/mean precipitation ratio would indicate that estimation over mid-latitudes is less successful than that over the tropics even in this parameter.

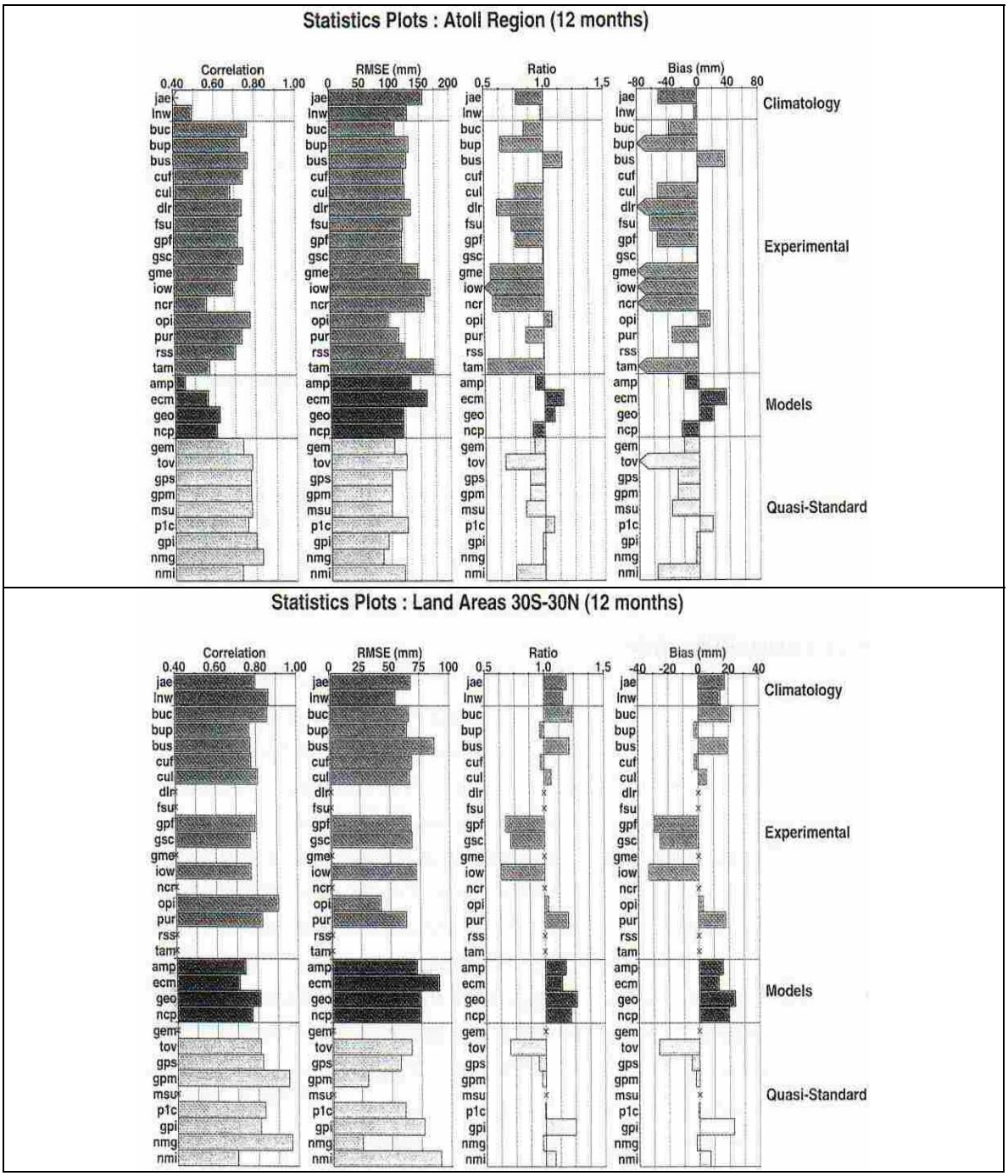


Figure 2.22 Statistical results for each of the products for monthly rain. A point on a bar indicates that the statistics is beyond the limits of the graph. (After Adler et al., 2001.)

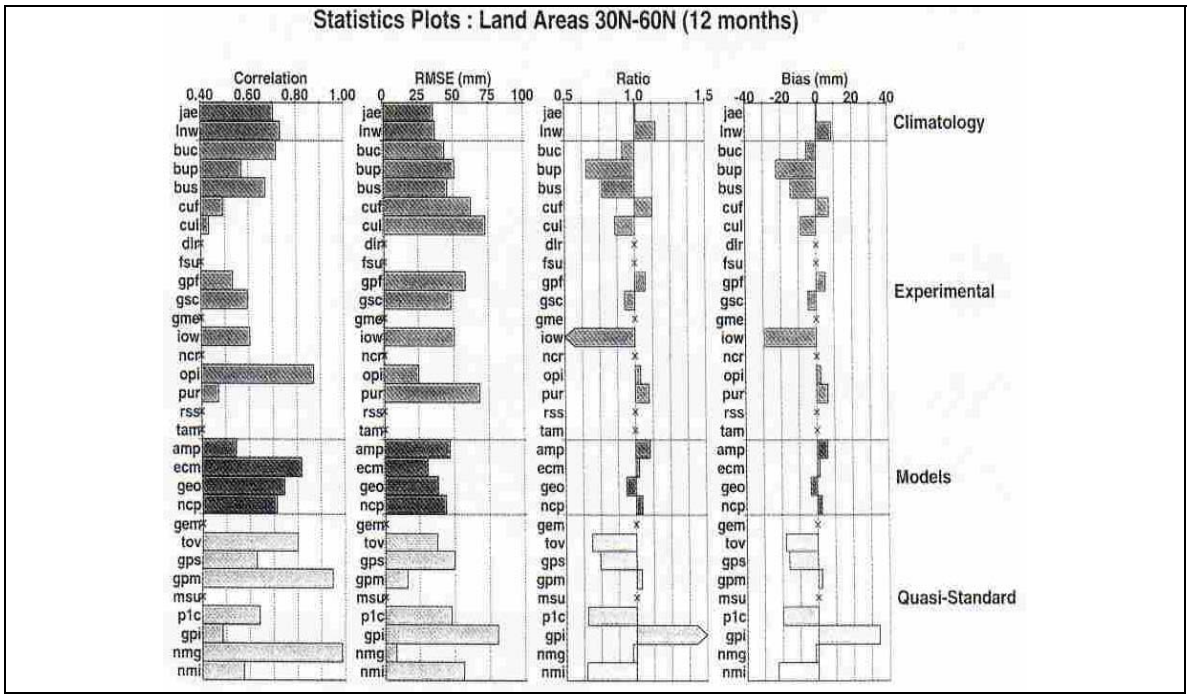


Figure 2.22 Continued.

2.3.3.5 Probabilistic Remote Sensing of Precipitation

Precipitation phenomena have been represented using probabilistic models with a variety of objectives, from projecting future precipitation distributions given present and past conditions, to interpolating rain gage measurements. This section explores only models that are connected to the remote sensing of precipitation. In this approach, the radiation observed in each pixel of a satellite image is not associated with a deterministic value, but with a probability distribution. Consequently, the resulting precipitation map is a random field, constituted of random variables with known, generally different, distributions. Further, some conditions are usually imposed on the second order moments in space and/or time of the precipitation random field. On average, the realizations of the random fields must satisfy both the single-pixel distributions and the constraints on the second order.

Not many authors have worked on this approach, probably for three main reasons:

1. Most efforts have concentrated on improving the deterministic model performance to make their estimates more credible;
2. The probabilistic treatment of remote sensing heavily increases the computation requirements of remote sensing techniques, which by themselves are already very computer intensive;
3. A probabilistic presentation of remote sensing results is more difficult to use and could decrease the, sometimes low, confidence in remote sensing products.

Although only indirectly related to the probabilistic remote sensing of precipitation, the paper by Bell (1987) proposed a stochastic model of precipitation that has inspired later models. The model was developed to evaluate sample strategies for remote sensing of precipitation from LEO satellites. The rainfall model has the following characteristics:

1. Precipitation is spatially homogeneous, temporally stationary, and isotropic;
2. At any grid point, precipitation is larger than zero just a fraction *f* of the time (precipitation intermittence). The rain intensity in rainy pixels is lognormally distributed.
3. The precipitation spatial correlation between any two points **x** and **y** is function of only the distance between the two points.
4. The correlated precipitation field is generated using a spectral method. That is, a gaussian field $Z(\mathbf{x})$ is generated by a Fourier series, the coefficients of which are Gaussian and uncorrelated variables.

$$Z(\mathbf{x}) = \sum_{\mathbf{k}} a_{\mathbf{k}} e^{i\mathbf{k}^T\mathbf{x}} \tag{2.10}$$

The field Z(x) is then transformed into the precipitation field.

5. Area-averaged rainfall rates have the time-lagged autocorrelation that depends on the size of the area. Larger areas have longer correlation times. This temporal autocorrelation is implemented by letting the coefficient a_k of 2.10 satisfy Markov equations in time:

$$a_k(t) = \exp(-\Delta t / \tau_k) a_k(t - \Delta t) + z_k(t) \qquad (2.11)$$

Where: Δt = time step

$z_k(t)$ = complex Gaussian white noise

$\tau_k = c_\tau(\pi/k)^{2/3}$, c_τ empirically obtained from lagged correlation analysis of mean areal precipitation

The model was calibrated using GATE data for the tropical areas of the Atlantic Ocean.

As mentioned in Section 2.3.3.2, Bell et al. (1990) used this model in a Monte Carlo study to characterize the error distribution of TRMM monthly precipitation estimates. Gremont (2002) included storm advection and used the model to study the attenuation in satellite communication signals due to precipitation.

Bell's model reproduces the statistics of a generic precipitation field with given rain/no-rain probability, conditional rain average and variance, and correlation structure, but it is not able to describe the statistic properties of specific weather conditions. Lanza (2000) describes a method to condition this generic precipitation field with precipitation rates measured by rain gages. His approach takes advantage of the linearity of the Fourier series and adjusts the value of the unconditionally simulated field to match the value of the conditioning nodes, while preserving the desired correlation structure and intermittence. In the case of one conditioning node, the convolution of the covariance

function with an impulse located at the conditioning node is added to the realization of the unconditional field. The amplitude of such an impulse is given by the difference between the unconditional value at the conditioning node and a known function of the measured rain rate and of the parameters of the unconditional distribution.

In the case of multiple conditioning nodes, the amplitudes of the single conditioning impulses are obtained by solving an appropriate linear system.

Bell's model and its variations are useful because they give a probabilistic framework for simulating precipitation at the spatial and temporal scales required by the remote sensing of precipitation. However, they cannot be used directly because they fail to include a spatial variation of the model parameters. Lanza's modifications may condition the random fields with a limited number of deterministic rainfall measurements, but they cannot accommodate the distributed and probabilistic information supplied by remote sensing.

A procedure closer to the problem at hand is that by Fiorucci et al. (2001). In their approach, the precipitation over an NxM grid is modeled as a lognormal multivariate distribution. The average and variance at each pixel k is related to IR radiation through the relation:

$$\mu_{R_k} = L_1\left(\frac{IR_k - C_0}{L_2} - 1\right)\exp(-\frac{IR_k - C_0}{L_2}) \qquad (2.12)$$
$$\sigma_{R_k} = (CV \cdot \mu_{R_k})^2$$

Where CV = 1.0

L_1, L_2, C_0 = empirical parameters.

Further, the covariance function is modeled as:

$$C_{LNR}(x_j, x_k) = \sigma_{LNR}(x_j)\sigma_{LNR}(x_k)\rho_{LNR}(|x_j - x_k|)$$
$$\rho_{LNR}(|x_j - x_k|) = \exp(-\frac{|x_j - x_k|}{\gamma} \quad (2.13)$$

Where $\sigma_{LNR}(x_k)$ = standard deviation of the $\ln(r(x_k))$ based on satellite information

$C_{LNR}(x_k,x_j)$ = covariance of the $\ln(r(x))$ between points x_k and x_j

$\rho_{LNR}(x_k,x_j)$ = correlation of the $\ln(r(x))$ between points x_k and x_j

γ = empirical parameter

Thus, the distribution of the NxM random field **Z** is described by:

$$f(Z) = \frac{1}{(2\pi)^{-MN/2}|\Sigma|^{1/2}}\exp(-\frac{(Z-\mu)^T\Sigma^{-1}(Z-\mu)}{2}$$
$$\mu = (\mu_{LNR_1}...\mu_{LNR_{NM}}) \quad (2.14)$$
$$\Sigma = \begin{bmatrix} C_{LNR_1,LNR_1} & . & . & C_{LNR_1,LNR_{NM}} \\ . & . & . & . \\ . & . & . & . \\ C_{LNR_{NM},LNR_1} & . & . & C_{LNR_{NM},LNR_{NM}} \end{bmatrix}$$

Fiorucci et al. (2001) do not use this formulation to obtain alternative realizations of the random field associated with an observed IR image. They instead use it to integrate gage measurements and satellite measurements in a Bayesian framework and obtain a single optimal estimate. If no gage measurements were available, then such optimal estimate of the precipitation would be just the expectation of the random field **Z**, which is vector μ in 2.14.

If precipitation is measured in the first G pixels, then each measurement can be considered as a realization of the random variables R_j, j=1, …, G, and the random vector **Z** can be partitioned into two subvectors $\mathbf{Z}_1$ and $\mathbf{Z}_2$:

$$\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2) = (Z_1, \ldots, Z_G, Z_{G+1}, \ldots, Z_{MN})$$
$$E(\mathbf{Z}) = \boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \tag{2.15}$$
$$\Sigma = \begin{bmatrix} C_{LNR_1, LNR_1} & . & . & C_{LNR_1, LNR_{NM}} \\ . & . & . & . \\ . & . & . & . \\ C_{LNR_{NM}, LNR_1} & . & . & C_{LNR_{NM}, LNR_{NM}} \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

The joint probability density function of $\mathbf{Z}_2$ conditional on the realization $\mathbf{z}_1$ of $\mathbf{Z}_1$ is given by:

$$f(Z_2 \mid z_1) = \frac{1}{(2\pi)^{-MN+G}} \left|\Sigma_{22.11}\right|^{-1/2} \exp(-\frac{(Z_2 - \mu_{2.1})^T \Sigma_{22.11}^{-1} (Z_2 - \mu_{2.1})}{2}$$
$$\mu_{2.1} = E(Z_2 \mid z_1) = \mu_{Z_2} + \beta_{2.1}(z_1 - \mu_{Z_1}) \tag{2.16}$$
$$\Sigma_{22.11} = \Sigma_{22} - \beta_{2.1}\Sigma_{12}$$
$$\beta_{2.1} = \Sigma_{21}\Sigma_{11}^{-1}$$

The conditional average $\boldsymbol{\mu}_{21}$ is used as the optimal estimate of the precipitation given the IR-image and the rain gage measurement $\mathbf{z}_1$.

Given that just the expected distribution of the random field is considered, no temporal correlation between consecutive random fields is required. Further, it is assumed that the average precipitation rate is always larger than zero. Non-rainy pixels are assigned a very small positive rain rate. Precipitation intermittency is obtained by screening out the negligible precipitation rates after the computation is complete. The application of this probabilistic model generates optimal conditional precipitation estimates and can evaluate the estimation variance at the single-pixel/single-time step level, but the evaluation the variance of the estimation over larger areas/longer periods is more difficult. However, equations 2.12-2.14, coupled with a method for generating the

realizations of a Gaussian field, could be used as a way to produce an ensemble of realizations of the precipitation field, which would more fully describe the uncertainty of precipitation estimates.

Seo (1998b) uses an approach similar to that of Fiorucci et al. (2001), but with two steps. In the first step the optimal rain/no-rain field is generated with a procedure similar to that of Fiorucci et al. (2001), but using a binomial random field in place of the longormal random field. In the second step, the optimal precipitation intensity is estimated supposing that it is normally distributed and conditioned by the satellite data, gage data, and the rain/no-rain field estimated in the first step.

Hossain et al., (2004) created a probabilistic model for assessing the impact of satellite PM rainfall retrieval and sampling errors on flood prediction for a medium-sized (~100 km^2) watershed. The approach they took was to generate a probabilistic distribution of the possible precipitation estimated by the satellite as a function of the gage precipitation. The simulated precipitation estimates were then fed to the rainfall/runoff model TOPMODEL to generate the flood prediction associated with the simulated satellite retrieval. Hossain et al. (2004) used this procedure in a Monte Carlo framework, systematically changing the satellite revisit period and time of the first passage. The ensemble of the model results quantified the uncertainty in flood prediction for a given revisit frequency. The authors modeled the simulated precipitation as composed of two components: a discrete rain/no-rain probability and a lognormally distributed precipitation intensity for the rainy pixels. The measurement error of the precipitation intensity was simulated as a lag-one autocorrelated process, but the small size of the considered basin allowed the use of only one satellite pixel to represent the area, thus avoiding dealing with the precipitation spatial correlation.

2.4 Summary

This chapter has described several aspects of precipitation and of the techniques used to quantify it using satellite images. One of the most relevant points of the chapter is that characteristics of storms are strongly dependent on the mechanism that lifts low-level moist air into the upper atmosphere, where it cools, generating clouds and precipitation. The type of lifting mechanisms prevalent in a region depends mainly on latitude, orography, and season. A very important lifting mechanism is convection, in which differential heating of the lower atmosphere rapidly lifts large quantities of moist air by buoyancy, generating intense precipitation. Different stages of the convective cells' lifecycle generate different precipitation regimes. More intense convective precipitation is prevalent during the early stages, whereas milder stratiform rain dominates the late stage of the convective cells' lifecycle, as buoyancy decays.

Section 2.2 shows that in many parts of the world, satellites represent the only economically or physically feasible way to measure precipitation. This consideration has prompted intense research on sensors and methodologies for remote sensing of precipitation. Indirect estimation schemes use the high-frequency IR images from geostationary satellites to determine the presence, type, and evolution of cloud systems. Precipitation is empirically estimated from this information (Section 2.3.3.1). The relation between IR data and precipitation is strong only at coarse temporal and spatial resolution. The addition of VIS and WV data for characterizing clouds makes the relation stronger, consenting to estimate precipitation at finer resolutions. Indirect precipitation estimation improves even more when factors such as storm history, seasonality, and orography are considered. However, such relations are subject to considerable variations in time and space, requiring dense networks of rain gages to optimize this relation in space and time.

Sensors in the microwave bands aboard LEO satellites provide direct estimates of instantaneous precipitation rates that are very reliable (Section 2.3.3.2). However,

estimation of precipitation over longer periods cannot be made using exclusively these data because of their poor temporal frequency, which is often less than one image every six hours.

Mixed schemes combine microwave images from LEO satellites with IR/VIS/WV images from geostationary satellites to exploit the precision of the former and the high temporal frequency of the latter (Section 2.3.3.3). The performance and flexibility of mixed schemes are generally better than those of either indirect or direct schemes alone and do not need to rely on extensive ground data for calibration. The major drawback is that they require online information from multiple sources.

Section 2.3.3.4 showed that remote sensing of precipitation is affected by substantial uncertainty. Improvements in satellite sensors and estimation methodologies improvements may reduce the level of this uncertainty, but cannot completely eliminate it. Furthermore, estimation uncertainty varies with time and spatial resolution, respectively ranging from less than an hour to days and months, and from a few to hundreds of square kilometers. Recognizing and quantifying such an uncertainty is important, but methodologies and procedures that deal with this subject are still in a development phase (Section 2.3.3.5).

CHAPTER 3

CONCEPTUAL FRAMEWORK

3.1 Conceptual Framework of the Probabilistic Estimation Procedure

The methodology presented herein belongs to the mixed scheme category in that precipitation estimation is based on IR/VIS images from geostationary satellites with the IR/VIS-precipitation rate relationship obtained from contemporaneous TRMM-PR and geostationary satellite images. The proposed methodology differs from most of the techniques presented in Section 2.3.3.3 in that it uses

- TRMM PR data rather than passive microwave radiation images;
- all available coincident TRMM PR and IR/VIS images to build the IR/VIS-precipitation rate relationship, not just data of the previous month;
- coincident PR and IR/VIS data from a 6ºx7º area, but distinguishes them according to orography features (lake/land and elevation).

TRMM PR estimates of instantaneous precipitation are much more reliable than passive microwave precipitation estimates, but they are approximately only 33% of the images supplied by the TRMM Microwave Imager. However, this deficit can be corrected by using all available coincident TRMM and IR/VIS images to build the IR/VIS-precipitation relation as opposed to monthly datasets. The use of such large data sets provides a more robust assessment of precipitation average spatial distribution and variability. This is essential for characterizing estimation uncertainty and for applying the procedure to periods for which microwave data are not available. In addition, larger data sets allow for regionalizing the IR/VIS-precipitation relationship based on orography

features rather than on latitude/longitude, leading to better precipitation estimates in heterogeneous terrains.

Sections 2.3.3.1 and 2.3.3.3 pointed out that the relation between observed IR radiation is affected by significant uncertainty at fine spatial and temporal scales. This uncertainty, however, is reduced when precipitation rates are also associated with other parameters, such as observed VIS/WV radiation, storm history, month, and pixel orography.

The methodology of this work utilizes a "look-up table" approach similar to those of King et al. (1995) and Kurino (1997). The table is indexed by orography, IR, VIS during daytime or IR-WV during nighttime, storm stage, and month. At each half-hour time slot, the precipitation over a pixel is determined as a function of the satellite observed IR/VIS/WV radiation in the pixel, orography, and season.

As mentioned in the introduction, one of the main goals of the methodology proposed in this research is to explicitly quantify the uncertainty associated with remote sensing of precipitation estimation over a variety of time and spatial resolutions. This is accomplished by treating the relation between satellite-observed radiation and precipitation in a probabilistic fashion as opposed to the more traditional deterministic way. Accordingly, an entire distribution of possible precipitation rates, instead of a single value, is associated with the observed radiation. Section 2.3.3.5 showed that when the precipitation's probability distribution has an analytical formulation, the spatial distribution of the optimal precipitation estimates and the associated variances could be mathematically derived using the appropriate summation formulas. There are, however, three problems with this approach:

1. It is not guaranteed that an adequate analytical model of the precipitation can be found;

2. As pointed out by Pardo-Iguzquiza et al. (2006), the precipitation is not likely normal at fine spatial and temporal scales. Thus, the precipitation mean and variance are not sufficient to fully describe the precipitation distribution.
3. The uncertainty in the output of hydrologic models using the estimated precipitation as input cannot be related to the mean and variance of the input, because the input/output relation is normally not analytical.

An alternative strategy is the ensemble technique, a convenient procedure when analytical forms of the probability density function are too difficult to derive. In this case, an ensemble of equally likely values is derived by randomly sampling a distribution of observed/simulated values. For example, Bell et al. (1990) used this model in a Monte Carlo study of the error distribution in TRMM monthly precipitation estimates. Walser et al., 2004, assessed the prediction capabilities of mesoscale numerical weather prediction (NWP) models by randomly perturbing the NWP initial conditions and observing the variability of the NWP output. Hossain et al. (2004), assessed the impact of satellite PM rainfall retrieval and sampling errors on flood prediction by randomly perturbing the timing and errors in the PM precipitation retrieval and by observing the variability of the output of a precipitation/runoff model using the precipitation estimates.

This study adopts the ensemble approach by associating each combination of input variables with the entire distribution of TRMM PR rain rates observed over a long period. Specifically, for each time-slot and pixel, an ensemble of equally likely precipitation values is obtained by randomly sampling the distribution of PR rates corresponding to the observed combination of IR, VIS, WV, storm stage, orography, and month. This approach is not limited by a particular probability model over the sample data. Further, it is easy to update and is relatively robust since outliers do not influence the entire distribution, but only a small number of cases. The disadvantages of this approach is that it requires keeping a lot of data in random computer memory and using a

Monte Carlo approach to derive the distribution of precipitation over a region/period. The estimation procedure for the precipitation over a single pixel can be delineated as follows:

1. The daily Meteosat IR time series (consisting of half-hour values) for a given pixel is processed by a neural network to determine the presence and evolution (storm stage) of convective storms (Figure 3.1 A).
2. At each timeslot t, the corresponding IR(t) and storm stage(t), defined in Step 1, identify a distribution of PR samples (Figure 3.1 B).
3. One estimate of daily precipitation is generated by randomly sampling the precipitation rates corresponding to the IR-stage combinations during the day and summing the resulting half-hour precipitation rates (Figure 3.1 C). The process is repeated N times resulting in N daily precipitation estimates (Figure 3.1 B).
4. To assess the accuracy and statistical consistency of the approach, the probabilistic position of the gage record with respect to the ensemble distribution is recorded (Figure 3.1 D). This value is used to determine various statistics such as the inter-quartile compliance rate, the 95% compliance rate, the Kolmogorof-Smirnov compliance statistic, among others.
5. The same approach is used to produce 10-day (dekad) and monthly precipitation estimates by extending the number of half-hour time series from 48 to 480, or to 48 times the number of days in the desired interval.
6. More sophisticated estimation procedures can be obtained by adding VIS(t)/WV(t), orography, and the calendar month in the set of indexing variables. A PR distribution similar to that of Figure 3.1 B will be identified by the appropriate combination of the input values in each time slot. Namely, for each member X of the ensemble:

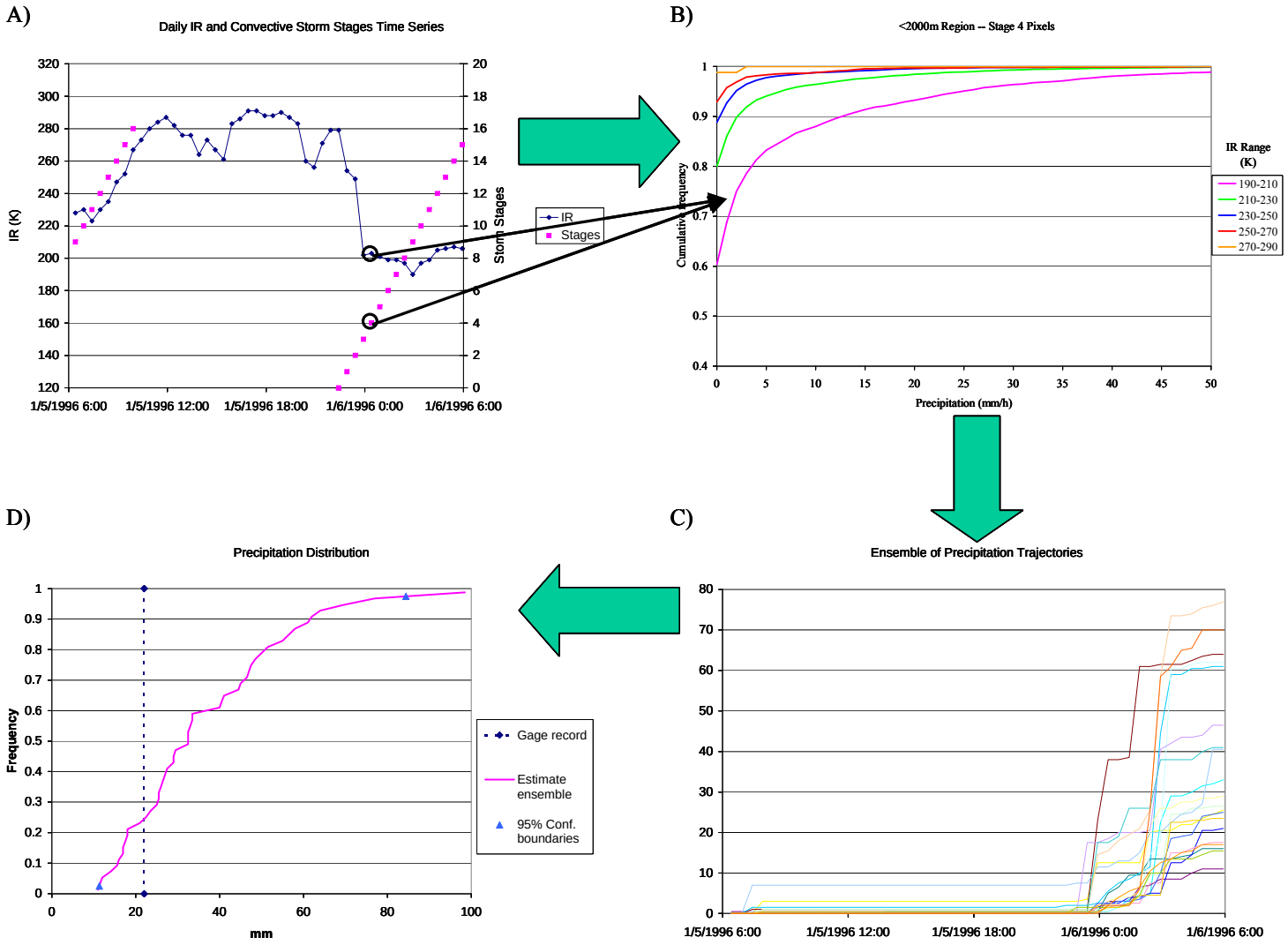


Figure 3.1 Schematic of the ensemble estimation of daily precipitation at the pixel level.

$$Precip(t) = f[IR(t), VIS(t)/WV(t), stage(t), orography, month, \varepsilon], \quad (3.1)$$

Where: Precip(t)= precipitation rate at time t for element X;

IR(t)= infrared radiation at time t;

VIS(t)= visible radiation at time t;

WV(t)= water vapor radiation at time t;

stage(t) =storm stage at time t;

ε = random error.

This approach can be extended over a region, by generating the random errors used for sampling the rain-rate distributions must be generated in a way that maintains the precipitation spatial correlation observed in the region. This topic is taken up in Chapter 8.

3.2 Development of the Estimation Procedure

The development of the estimation procedure follows a step-by-step approach, starting from the single components at the single pixel/single time-step level and adding complexity as required by the expansion of the spatial and temporal domains. This approach has two advantages:

1. It allows a very complex process to be broken into smaller sub-processes that are easier to treat;
2. It allows an optimal use of all available data for faster and more complete calibration/verification of single components and sub-models using large datasets, while restricting the test of the complete model, which is more demanding in terms of computation time and testing data, to a smaller dataset.

The major steps in the development are briefly described in the following subsections and analyzed thoroughly in later chapters.

3.2.1 Identification of Convective Storms

Section 2.1 showed that convective storms feature three distinct phases – developing, mature, and dissipating – each with a distinct rain regime. Using specific relations between cloud characteristics and precipitation for these different phases of the convective storm and for different types of storms should yield better precipitation estimates. Such an approach to precipitation estimation involves two steps:

1. Detecting both the onset and termination of the convective storm;
2. Determining the relation between precipitation and cloud characteristics for different parts of the storm.

The detection of the occurrence of a convective storm over a pixel is done by a Multilayer Feed-Forward Neural Network (MFFNN) specifically trained to recognize a catalog of 3-hour IR patterns associated with the onset of typical convective patterns. The convective patterns were selected among those corresponding to pixels categorized as convective in a contemporaneous TRMM PR swath. The non-convective patterns were chosen from among the patterns corresponding to pixels categorized as non-convective by TRMM.

Because the MFNN is trained to recognize IR-patterns corresponding to the onset of a convective storm, it also indirectly identifies the storm temporal evolution.

This procedure is illustrated in a more comprehensive way in Chapter 5.

3.2.2 Empirical Versus Analytical Distributions

Distributions of precipitation rates are associated with all combination of IR, Stage, VIS/WV, month, and orography by sampling a multi-year database of contemporaneous TRMM and IR/VIS/WV data. The estimation of the precipitation is accomplished both by using the sample distributions (empirical distributions) and directly by fitting an analytical model to the sample distributions (analytical distributions). The model used is the one proposed by Bell (1987), in which the probability of no-rain is given a discrete value, while positive rain intensities are described by a lognormal distribution:

$$f_R(z) = \begin{cases} 0 & z < 0 \\ P(0) & z = 0 \\ P(0) + N(\ln(z), \mu_{LNR}, \sigma_{LNR}) * (1 - P(0)) & z > 0 \end{cases} \qquad (3.2)$$

Where: $f_R(z)$ = probability density function of precipitation R;

$P(0)$ = probability of no rain;

$N(\ln(z), \mu_{LNR}, \sigma_{LNR})$ = lognormal distribution of positive rain rates;

Differently from Bell's approach, however, μ_{LNR} and σ_{LNR} vary with the observed IR, VIS/WV, storm stage, month, and orography.

This procedure is illustrated in a more comprehensive way in Chapter 6.

3.2.3 Single-Pixel Analysis

The ability of the remote sensing procedure to reproduce precipitation patterns is assessed by comparing precipitation estimates and rain gage measurements over a set of single pixels. Standard statistics such as bias, correlation, rank correlation, and mean average error are used to assess the capability of the average precipitation to track the

single gage records. The capability of the model to reproduce the data variability is evaluated using the inter-quartile compliance rate, the 95% compliance rate, and the Kolmogorof-Smirnov compliance statistic. The procedure performances were optimized by comparing estimation results and measurements coming from a multiyear database including a large number of rain gages. Estimations were also verified against a second set of precipitation records for gages not included in the calibration dataset. Precipitation temporal correlation is also introduced during the single-pixel analysis. These procedures are illustrated in a more comprehensive way in Chapter 7.

3.2.4 Multiple-Pixel Analysis

The single-pixel analysis assumes that precipitation is spatially uncorrelated. This approach is convenient because it allows calibration/validation of the average model performance by considering only the pixels containing rain gages, generally a very small subset of the region for which the procedure is applied. Ignoring the precipitation spatial correlation, however, does not permit proper estimation of the precipitation uncertainty over larger areas, which are normally considered in water resources management. Thus, the precipitation estimation routine must be modified to account for the spatial correlation of rainfall. Implementation of the spatial correlation regards both the empirical-distribution and the analytical-distribution models.

Multi-pixel performances of the model, with and without spatial correlation, were evaluated using indicators similar to those described in section 3.2.3, but considering the average precipitation over sets of neighboring gages. The Multi-pixel approach is illustrated in a more comprehensive way in Chapter 8.

CHAPTER 4

AVAILABLE DATA

4.1 The Hydrology of the Lake Victoria

The proposed precipitation estimation methodology has been evaluated in an area surrounding Lake Victoria. More specifically, this area extends from 29ºE to 36ºE and from 5ºS to 3ºN (Figure 4.1 and Figure 4.2). Lake Victoria covers roughly the central 10% of this region. This region is flanked by high mountains, with Mount Elgon (4155 m) in the east side and Mount Ruwenzori (5109 m) in the west.

The climate of the region is equatorial, but elevation and lake influence contribute to moderate temperatures all year round. The lowlands in the southern side of the area are considerably drier than the rest of the basin.

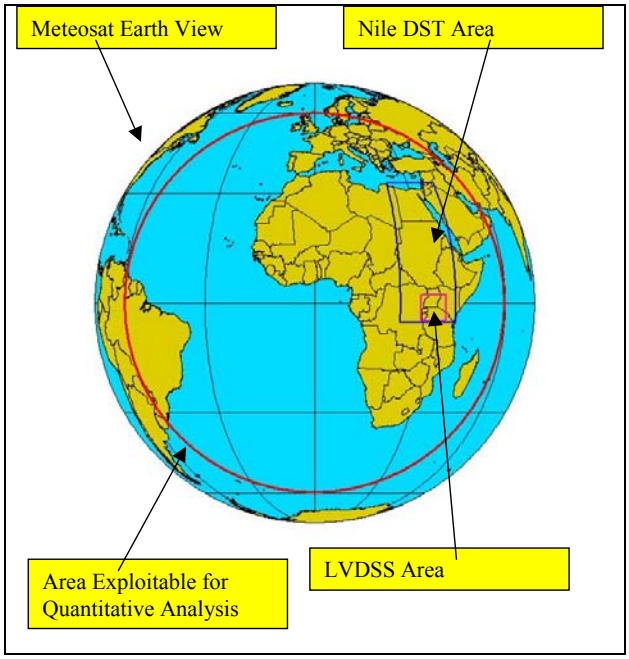


Figure 4.1 Full Disk View of the Meteosat Satellites.

Precipitation is driven mainly by the migration of the Inter Tropical Convergence Zone with the related northeast and southeast monsoons. Lake Victoria is large enough to partly modify the general circulation by creating a permanent pressure low attracting moisture from the Congo rainforest in the west. Further, it creates local precipitation patterns similar to those created by land-sea interaction. The high mountains in the basin also influence the climate by creating windward and leeward regions.

This section illustrates the satellite and rain gage data available to this project as well as data available for future modeling developments. Data processing for quality control is also described.

4.2 Rain Gage Data

4.2.1 Characteristics of Rain Gage Data

The precipitation estimation procedure has been evaluated using the December 2002 version of the FAO Nile Data Base (NBD-Dec02-2) and some additional data from the Kenya Ministry of Water, Tanzania Ministry of Water, and the Egyptian Nile Forecast Center. This data were available to GWRI/Georgia Tech as part of the development of the Nile DST and other projects.

4.2.2 Rain Gage Data Processing

Rain gage data have undergone extensive examination to ensure the quality of the records. The main quality control steps were as follows:

1. Rain gage coordinates have been verified against GIS maps and WMO or national identification codes;
2. Duplicate and ambiguous precipitation records have been eliminated;
3. Monthly sequences of daily precipitation with less than 90% records have been dismissed;

4. For each rain gage, the average precipitation during each month of 1996, 1997, and 1998 has been compared against the 15-year average precipitation. Months with precipitation outside the normal range or months with insufficient records were compared against measurement at neighboring stations;
5. Correlation in 10-day precipitation between neighboring stations was checked;
6. Only stations featuring at least 30 months of valid precipitation records in the 1996-1998 or 20 months of valid precipitation records in the 1996-1997 periods were retained;
7. Differences in the correlation between daily precipitation and satellite estimates at neighboring stations were investigated;

The resulting data records that passed the previous tests are reported in Table 4.1 and in Figure 3.3.1. The rain gage density in the basin fluctuates significantly making the examination of the gage records particularly difficult. Therefore, it is likely that despite all quality control efforts, rain gage data still suffer from a higher percentage of error than similar data from other parts of the world.

Table 4.1 Available rain gage data in the Lake Victoria region.

Period	Stations with temporal coverage of at least 83%	Source
01/1996-12/1997	Lake Victoria basin - 98	FAO
01/1996-12/1998	Lake Victoria basin - 61	FAO
01/1998-12/1998	Lake Victoria basin - 48	FAO

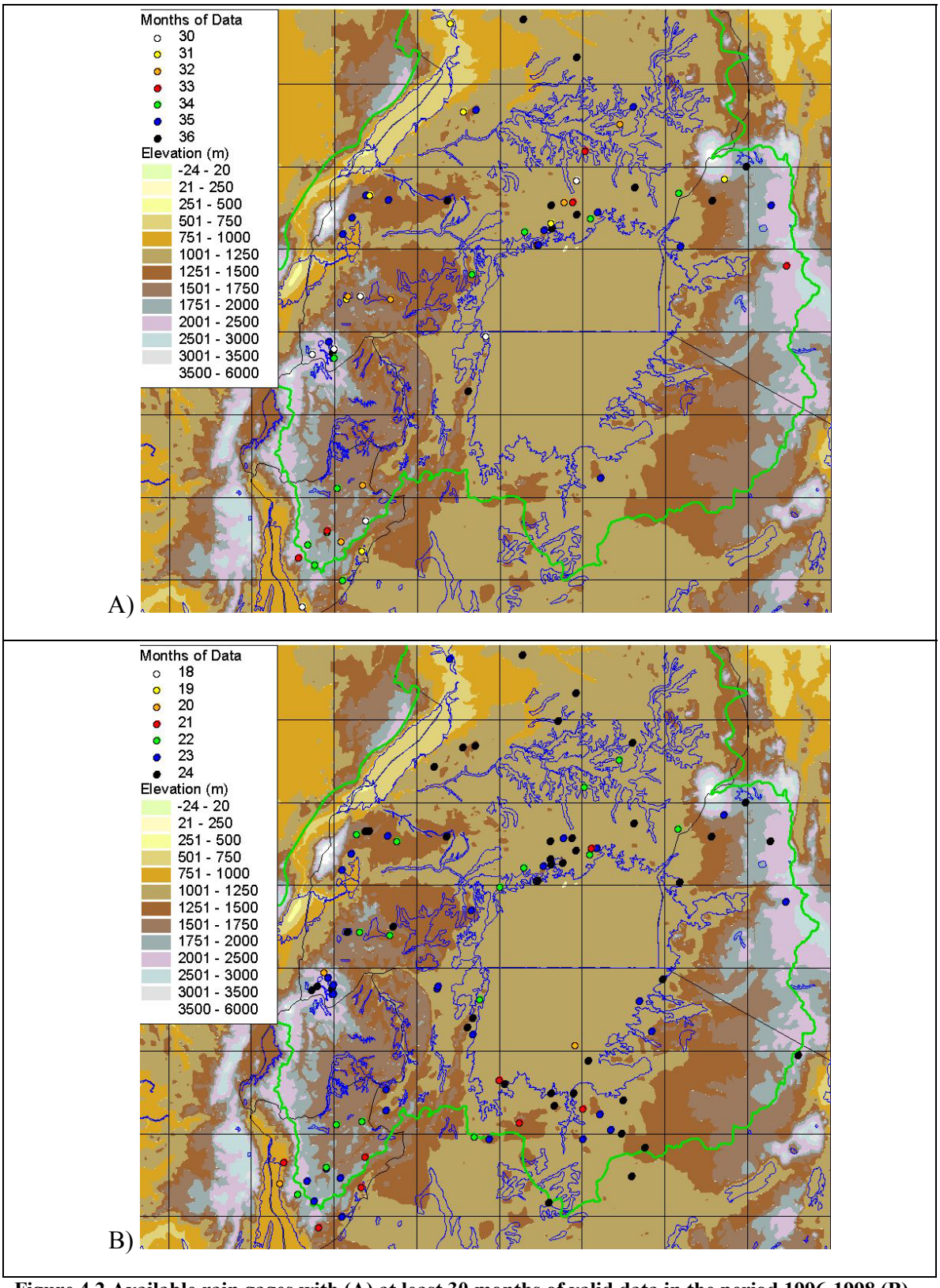


Figure 4.2 Available rain gages with (A) at least 30 months of valid data in the period 1996-1998 (B) at least 20 months of valid data in the period 1996-1997.

4.3 Satellite Data

4.3.1 Meteosat Data

The general characteristics of the Meteosat images have been extensively described in Section 2.3.2.1. Four different Meteosat satellites, Meteosat-4 through Meteosat-7, have provided digital images over the period 1992-2003. While the general characteristics of images have remained those described in Section 2.3.2.1, each satellite's sensors require specific digital count – radiance relations. Further, performances of sensors vary during their operational life because of sensor degradation and environmental influences, requiring a continuous recalibration of the count-radiance relationship. The satellite-specific sensor count-radiance relation and the calibration coefficients can be retrieved from the Eumetsat web site (www.eumetsat.de). Some calibration coefficients are also embedded in the satellite broadcasted digital images as auxiliary parameters. The quality of the image (i.e., the noise level) is also different from satellite to satellite.

Meteosat digital images were available to this project in three different formats, namely B.U.R.S.L. Autosat 3.1 (FAO/UN – Egypt images before December 1996), B.U.R.S.L. Autosat Block 5 (FAO/UN – Egypt images after December 1996 and FAO/UN – Uganda images), and Eumetsat OpenMTP. All three formats hold much more information than needed for this study, dramatically increasing storage requirements. Further, extracting the data from these files is a cumbersome process that easily fails if the digital files are corrupted. It also requires merging data files in different formats for producing a continuous sequence of images. For these reasons the digital images have been re-sampled over a regular rectangular grid with 0.05ºx0.05º resolution and transformed into a nimbler intermediate format called NileDST Grid Format. (NileDST stands for the Nile Decision Support System that has been developed by the

Georgia Water Resources Institute at Georgia Tech, Georgakakos, 2004.) The translation process included the following steps:

1. Elimination of missing and noisy lines;
2. Conversion of digital counts into blackbody temperatures (IR and WV channels) and albedo (VIS channel) with correction of the solar-angle effect for the latter.
3. Filtering of clearly invalid values;
4. Filtering of pixels outside the range $AVG_{5x5} - 3 \times STDEV_{5x5} <$ pixel value $< AVG_{5x5} + 3 \times STDEV_{5x5}$, where AVG_{5x5} is the channel average over a five pixels by five pixels square centered on the pixel of interest, and $STDEV_{5x5}$ is its standard deviation;
5. Detection and elimination of corrupted, spatially or temporally displaced, and duplicated images;
6. Correction of the displacement error.

Table 4.2 Available Meteosat data.

Period	Spatial and temporal coverage of images	Source
06/1992-12/1994	South of 22N	FAO – Egypt
01/1995-12/1995	South of 22N – 96%	FAO – Egypt + Eumetsat
01/1996-12/1996	South of 22N – 98%	FAO – Egypt + Eumetsat
01/1997-12/1997	Nile Basin – 98%	FAO – Egypt + Eumetsat
01/1998-09/17/1998	Nile Basin - Complete	Eumetsat
09/18/1998-12/1998	South of 10N – Complete	FAO Entebbe + Eumetsat
01/1999	South of 5N– Complete	Eumetsat
02/1999-25/05/1999	Nile Basin - Complete	Eumetsat
06/05/1999-04/2000	Nile Basin – 88% coverage	FAO – Egypt
05/2000-07/2000	None	
08/2000-14/09/2001	Nile Basin – 83%	FAO – Entebbe
15/09/2001-12/29/2002	Nile Basin – 83%	FAO – Entebbe

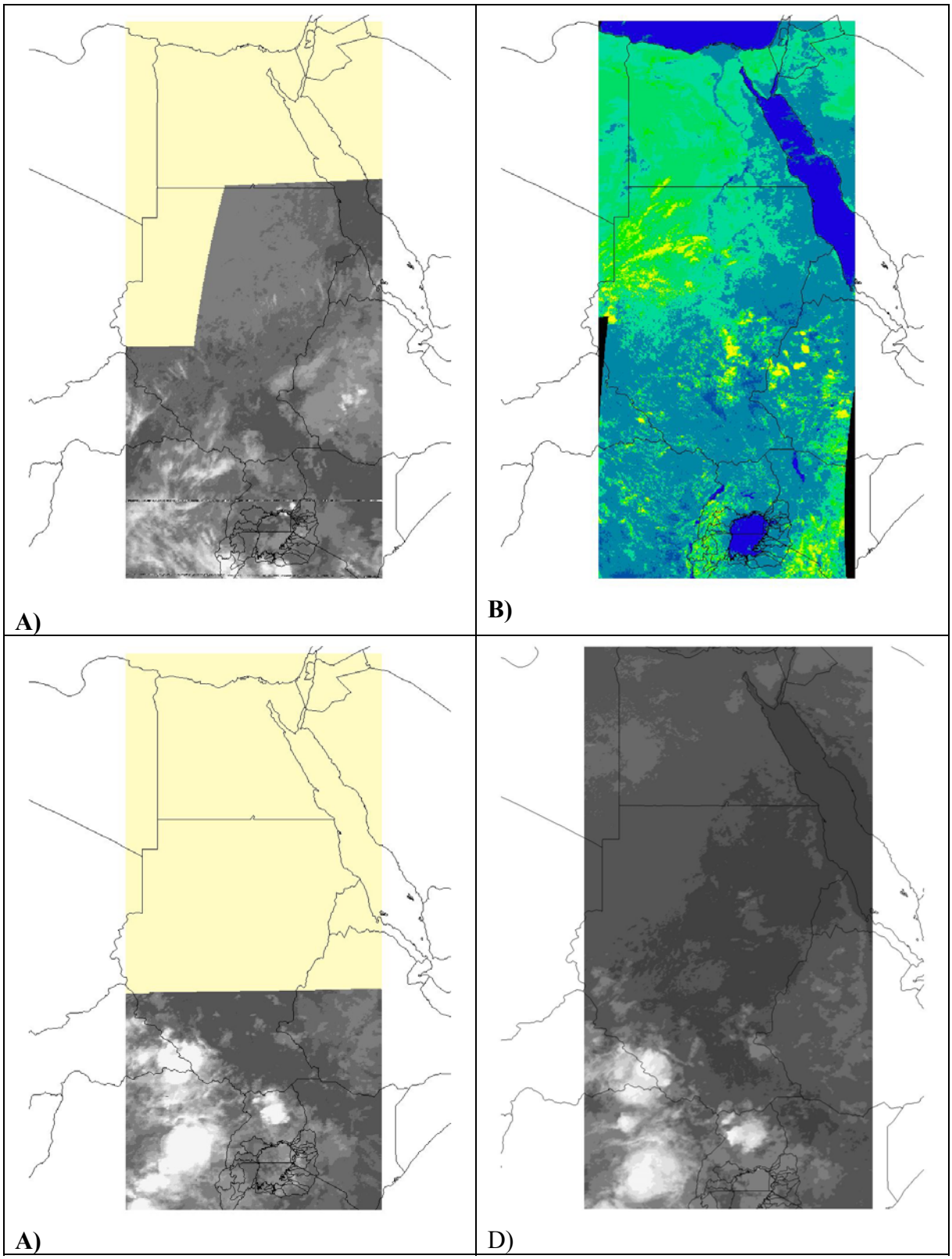


Figure 4.3 A) Example of FAO-Egypt IR image until 1996. A line of noise slightly North of Lake Victoria is maintained for illustrative purposes. B) Example of FAO-Egypt VIS image after 1996. C). Example of FAO-Entebbe IR image in 1998. D) Example of Eumetsat IR image.

Satellite digital images in 1996-1997 were frequently affected by one or more lines of noise like the one shown in Figure 4.3. Quality of the remaining files range from good to optimal in the case of the OpenMTP files.

4.3.2 TRMM Data

The general characteristics of the TRMM images have been extensively described in Section 2.3.2.2. The data used for this study are products "2A23 – PR Qualitative" and "2A25 – PR Profile" data, version 5. The former product first determines whether each range bin is above or below the ground surface and if the corresponding radar reflectivity is reliable. Then, the presence of rain is assessed as "possible" (the return power is larger than the 90%-tile noise) or "certain" (the return power exceeds the "possible" noise level by more than three times the noise standard deviation).

PR assesses the rain type by merging the results of two different methods for classifying rain type: the H-method based on the horizontal reflectivity distribution (Steiner et al., 1995) and the V-method based on the reflectivity vertical profile. The final precipitation classification is obtained by merging the outcomes of the H-Method and V-Method with other information in the bright band (BB), resulting in a complex categorization of the rain type. The project, however, retained only the major precipitation classes (stratiform, convective, and "other") by using the following scheme (TSDIS, 2004)

$$\begin{aligned}\text{(merged) rainType[i] / 100} = \;& 1\text{: stratiform;} \\ & 2\text{: convective;} \\ & 3\text{: other.}\end{aligned} \qquad (4.1)$$

Smearing of BB near the antenna scan edges seriously affects the BB detection, which is about 80% for antenna scan angles in the interval ±7° from nadir, but only about

20% at the swath edges. Further, the rain/no-rain discrimination seems also to be dependent on the angle-bin. To alleviate the impact of side-lobe clutter, the project does not use the first and last two pixels of each PR swath.

TRMM data have been re-sampled along the 0.05ºx0.05º NileDST grid for matching with contemporaneous Meteosat data. When two or more TRMM pixels fall within the same NileDST pixel the resulting rain is classified as "entirely convective", "entirely stratiform", "entirely mixed", or "no-rain" if the original TRMM data were of the same type. If they were of different types, the resulting NileDST pixel is classified as "partially convective", "partially stratiform", or "partially mixed" (Figure 4.4).

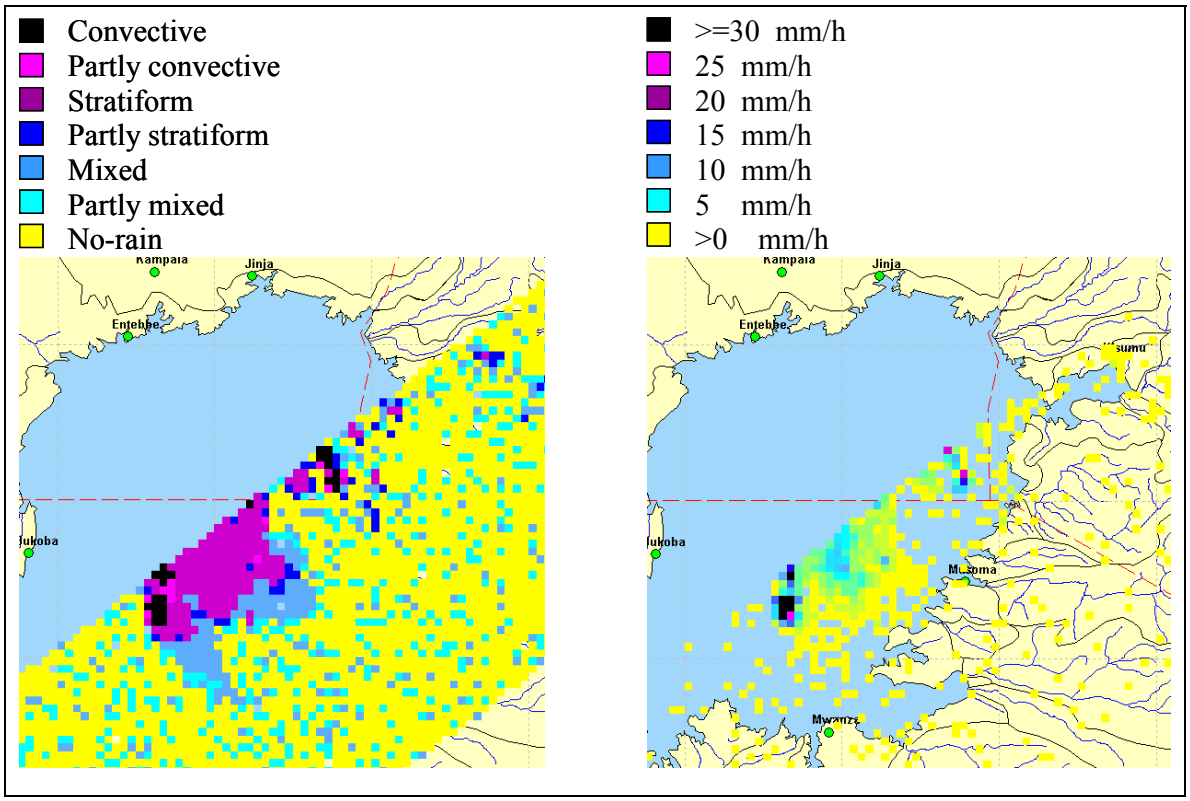


Figure 4.4 Example of rain type and near surface rain-rate information provided by TRMM.

Product "2A25 – PR Profile" reports the precipitation rate as well as its reliability for each beam of the PR swath over 80 vertical levels. It also includes additional information relevant to the precipitation rate retrieval procedure.

It was not possible to use precipitation rates at a predetermined elevation or the average rain between 2 and 4 kilometers, two TRMM products that are frequently used for precipitation estimates, because of the large variation in elevation in the Lake Victoria basin makes them either applicable only to a limited portion of the area or too far from the surface. Near surface rain, on the other hand, is available for all pixels of the basin, thus allowing a complete coverage of all elevation bands. It also accounts for precipitation evaporation along raindrop trajectories, an issue that may be relevant in the drier parts of the basin. On the other hand, near surface rain is likely to be noisier than precipitation rates averaged over several vertical levels and more strongly affected by ground clutter. A basin-wide comparison between precipitation rates near the surface and at 2 kilometers altitude suggested that near surface rates below 0.7 mm h^{-1}r are very likely to be noise and were thus eliminated.

As with the rain type product, 2A25 pixels have been re-sampled to fit the 0.05ºx0.05º NileDST grid by averaging the precipitation of multiple TRMM pixels falling within the same NileDST pixel.

Table 4.3 Available TRMM data.

Period	Spatial and temporal coverage of images	Source
01/1996-12/1996	None	
01/1997-12/1997	None	
01/1998-12/1999	-5N÷5N, 28E÷38E 816 valid TRMM passages	University of Connecticut
12/1997-Present	Nile Basin Acquisition in process	NASA

CHAPTER 5

IDENTIFICATION OF CONVECTIVE STORMS

5.1 Chapter Scope

Most precipitation estimating procedures using infrared and visible images from geostationary satellites have an "instantaneous" nature in the sense that the estimation of the precipitation at a certain time is based only on the satellite images taken at that particular time. Some procedures recognize the importance of the temporal change in cloud characteristics and base their estimates on pairs of consecutive satellite images (Scofield, 1987; Vincente et al., 1998; Bellerby et al., 2000). However, these procedures do not take into consideration the fact that convective storms generally feature three distinct phases – developing, mature, and dissipating – each with a distinct rain regime. Using specific relations between cloud characteristics and precipitation for these different phases of the convective storm and for different types of storms should yield better precipitation estimates. Such an approach to precipitation estimation involves two steps:

1. Detecting the onset and termination of the convective storm;
2. Determining the relation between precipitation and cloud characteristics for different parts of the storm.

This chapter is organized as follows: the beginning of Section 5.2 describes the typical IR patterns associated with convective storms and outlines the procedure for identifying convective storms by using an artificial neural network. Section 5.3 explains the procedure used to characterize the temporal evolution of convective storms. Finally, Section 5.4 summarizes the main points covered in the chapter.

5.2 Detecting Convective-Storm Occurrence at the Pixel Level

Figure 5.1 shows the visible and infrared signals associated with a typical tropical convective storm at the pixel level. Initially, clouds are low (high IR temperature) and relatively thin (low albedo). As the convective storm matures, the clouds get thicker (rising albedo), and the cloud top approaches the tropopause becoming colder (falling IR temperature). As the convective cell dissipates, the albedo decreases and the IR temperature increases.

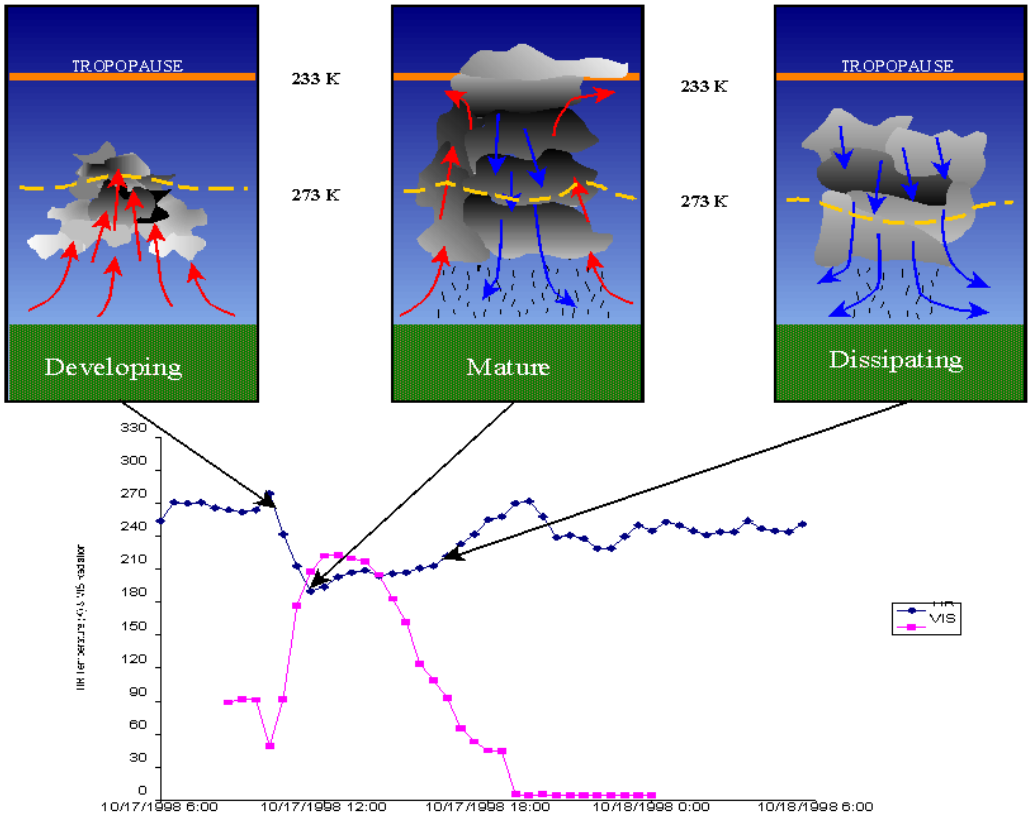


Figure 5.1 Infrared (Diamonds) and Visible (Squares) patterns during a typical convective storm.

Detecting the occurrence of a convective storm over a pixel means identifying the presence of a pattern that can be associated with this phenomenon, such as that in Figure

5.1, in the time series of satellite infrared and visible data for that pixel. This task is made more complex by the fact that highly variable factors, such as storm duration, intensity, and antecedent condition, confer a large variability to these "signatures". The methods used for recognizing the presence of convective storms vary from generic simple algorithms, like IR-thresholding (Richards & Arkin, 1981) and IR/VIS look-up tables (King et al., 1995) to complex approaches involving, for example, the IR radiation and its time derivative (Vincente et al., 1998) or the IR/VIS texture (Hsu et al., 1997, 1999). Georgakakos et al. (2000) tested several methods for identifying the presence of such typical patterns. The best results in terms of effectiveness and computing requirements were obtained by adopting a Multilayer Feed-Forward Neural Network (MFFNN) specifically trained to recognize a catalog of 6-hour long IR and VIS patterns associated with typical convective patterns. The convective patterns were selected by determining storm centers, or cores, where numerous adjacent patterns are identical and similar to that of Figure 5.1. Farther away from these storm cores, patterns were identified as non-core, or transitional, when they began to lose that typical convective shape. As a consequence of this approach, the number and variety of patterns used for training were limited. This work improves the approach of Georgakakos et al. (2000) in the following aspects:

1. The IR patterns selected for training the MFFNN are not chosen according to their resemblance to an idealized model, but because they correspond to pixels categorized as convective in a contemporaneous TRMM PR swath. The non-convective patterns are chosen among the patterns corresponding to pixels categorized as non-convective by TRMM.
2. TRMM PR rain type information is also used to validate the MFFNN ability to discriminate between convective and non-convective patterns.

3. The neural network is trained to recognize only IR patterns, since the daily variation of the VIS signal, especially that transmitted by Meteosat satellites, makes long sequences of VIS data difficult to use.
4. The IR patterns are composed of between five and eight IR values around the onset of the developing phase of the storm. Shorter IR patterns limit the computation time, reduce the negative effect of missing slots, and are less sensitive to storm duration variability.
5. The explored MFFNNs architectures include both three-layer and four-layer MFFNNs with several combinations of nodes.

Artificial Neural Networks (ANN) are mathematical devices designed to mimic the information-processing patterns of the biological neural networks in the brain. Although several different types of ANN have been developed, they all feature some common building blocks (ASCEE, 2000a):

1. ANNs are composed of elementary processing units (called nodes or neurons) connected to each other by unidirectional arcs.
2. Each arc has an associated weight that represents the strength of the connection between the two connected nodes.
3. The output of each node is fed to the nodes to which it connects after being multiplied by the weight of the connection.
4. The output of each node is the result of a (generally non-linear) transformation of the net sum of all the weighted inputs received from the other nodes. This transformation is called the activation function.

Figure 5.2 delineates the structure of the MFFNN, probably the most widely used ANN, adopted for convective storm detection. The nodes of MFFNNs are usually

arranged in layers, starting from the input layer, which receives information from the external world, and ending with the output layer, which communicates the result of the overall processing to the external world. One or more layers of nodes (hidden layers) are placed between these two layers. Information passes only from nodes of one layer to the nodes of the following one. No node sends information to other nodes that are in the same layer or in previous layers.

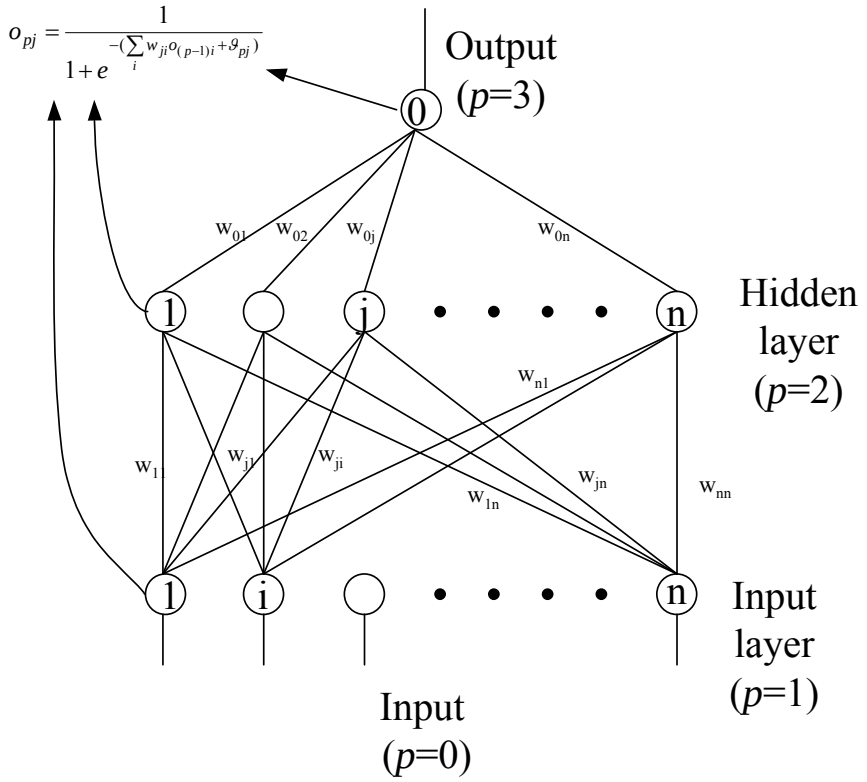


Figure 5.2 Structure of a three-layer feed-forward neural network.

The weight of the arcs and the bias of the activation functions can be calibrated to replicate a desired input-output relation, through the following iterative process, knows as ANN training:

1. Weights and biases of the ANN are initially set to some initial value.
2. The resulting ANN is used to compute the responses to a set of input patterns.
3. These responses are compared to the desired outputs and the weights and biases are adjusted to reduce the discrepancy between the two outputs.
4. Steps 2 and 3 are repeated until the discrepancy between ANN output and desired output is below certain limits.

The algorithm for updating the MFFNN weights and biases adopted in this study is the Back-Propagation Algorithm, also known as the Modified Delta Rule (ASCE, 2000a). A detailed description of neural networks and their calibration and application in hydrology can be found in (ASCEE, 2000a, 2000b).

The neural networks tested for convective pattern recognition have the following features:

- Three or four-layer MFFNN with sigmoidal activation functions;
- Number *n* of input values varying between five and eight;
- Number of nodes in the first hidden layer (the one closer to the input layer) set to *n*, 2**n*, 3**n*, 4**n*, or 5**n*;
- Number of nodes in the second hidden layer (if present) set to *n*;
- Output value close to one if a convective pattern is present in the input sequence, close to zero otherwise.

ANN training is done offline and does not interfere with operational convective storm identification. Namely, for each pixel (*i*,*j*) and time slot *t*, the ANN determines if a convective pattern is present in the sequence IR(i,j,t),…,IR(i,j,t+n) (Figure 5.3).

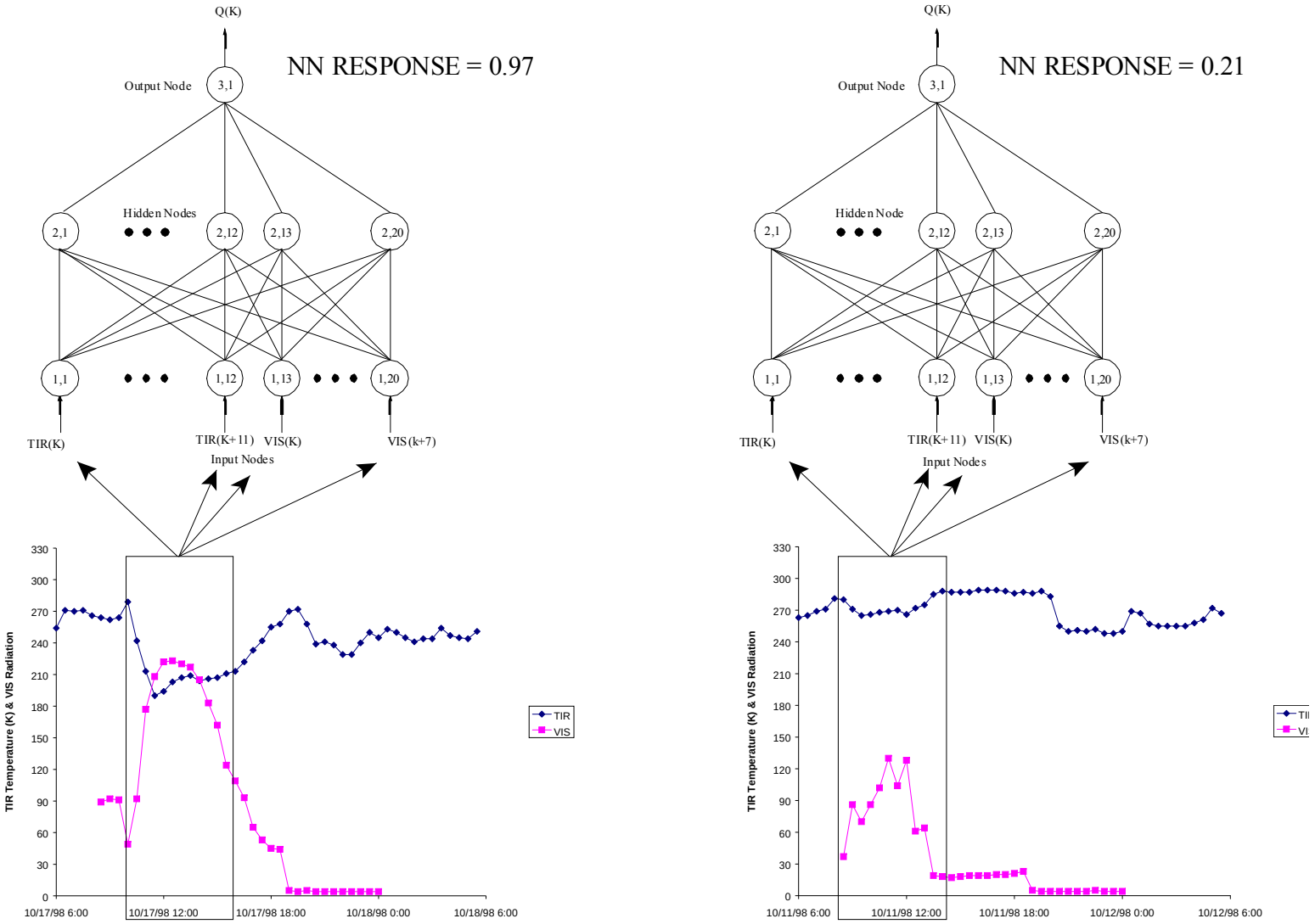


Figure 5.3 Schematic of MFFNN application for convective storm identification.

5.2.1 Training the Neural Network for Storm Identification

As mentioned in the previous section, the MFFNN is trained to recognize a set of convective and non-convective IR patterns (training set). A key issue in this process is identifying when to stop the training. If a neural network is under-trained, it will be able to identify only the dominant features of the patterns, resulting in large errors. On the other hand, if over-trained, the network essentially memorizes the specific patterns in the training set, but cannot identify other similar signals. In view of this, the approach adopted herein was to train the network on a first data set, and to determine when to stop training based on its predictive performance over a second data set (verification data set).

5.2.1.1 Selecting the Convective Patterns for the Training/Verification Sets

TRMM PR rain type data from the first ten days of each month of 1998 for a total of 1081 patterns were used to select the convective pattern candidates for the MFFNN training and verification sets. Past work with TRMM data in the Lake Victoria region suggested that the IR patterns associated with convective storms could be clustered into three clearly distinct shapes (Figure 5.4):

Deep and steep convection

The IR traces belonging to this class (Figure 5.4 A) are characterized by a strong and rapid IR dip (minimum IR below 230 ºK; dIR/dt of at least –26 ºK hr^{-1}). In Figure 5.4 A, the solid purple line marks the borderline with shallow convection, while the green and triangle line marks the borderline with deep, but slowly building, convection. This class accounts for 70.2% of the patterns, and features an average rain rate of 27 mm hr^{-1}.

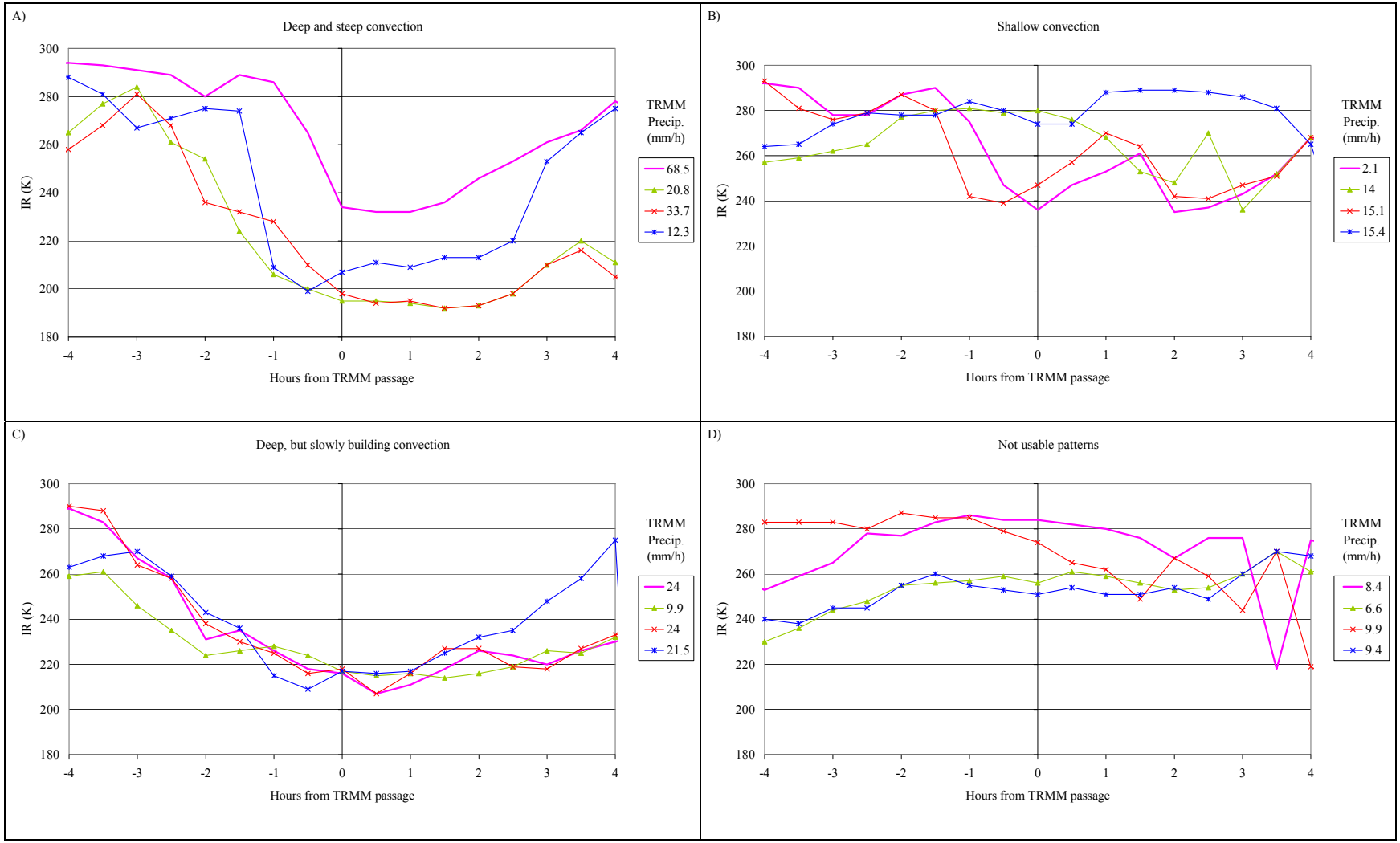


Figure 5.4 Convective pattern classification.

Shallow convection

The IR traces belonging to this class (Figure 5.4 B) are characterized by a shallow convection (minimum IR above 230 ºK) and are frequently found over mountainous areas. This category accounts for 14.2% of the patterns and features an average rain rate of 17.4 mm hr^{-1}.

Deep, but slowly building convection

These IR traces (Figure 5.4 C) display strong, but gradual, IR dips (minimum IR below 230 ºK; dIR/dt of less than –20 ºK hr^{-1}) and are frequently found over the lake surface. This class accounts for 13.4% of all the patterns and has an average PR surface rain of 22.8 mm hr^{-1}.

A small fraction of the patterns (2.2%) do not fall into any of these three categories nor do they show an alternative common behavior (Figure 5.4 D).

Mixing these distinct patterns in the same training/verification sets would have diminished the sharpness of the ANN. Since the large majority of the convective patterns fall in the first category, which features the largest average rain rate, it was decided to train the neural network to recognize only this type of convective patterns. Further, in order to emphasize the similarity in the patterns, the convective patterns were aligned around the time at which the IR dip reaches IR=240 ºK (Figure 5.6).

5.2.1.2 Selecting the Non-Convective Patterns for the Training/Verification Sets

A set of patterns corresponding to non-convective patterns must be included in the training set. These patterns (Figure 5.5) include clear sky (solid purple line) and shallow convective patterns (green line with triangles). Since the purpose of the neural network is to mark the onset of a convective storm, the set also incorporates IR traces corresponding

to the period antecedent to the storm (red line with x), its mature stage (blue line with asterisks), and its dissipating phase (dashed dark blue line).

TRMM PR Rain Type maps from the first ten days of each month of 1998 were used to (1) identify non-convective pixels that were at a distance of at least one or two pixels from the nearest convective pixel, and (2) to determine the direction in which storms were moving in order to select pixels at their antecedent, mature, and dissipating phases. With these additions, the training and verification sets each consisted of 379 convective and 379 non-convective patterns and listed alternating elements of the two types to achieve a balance between the convective and non-convective patterns.

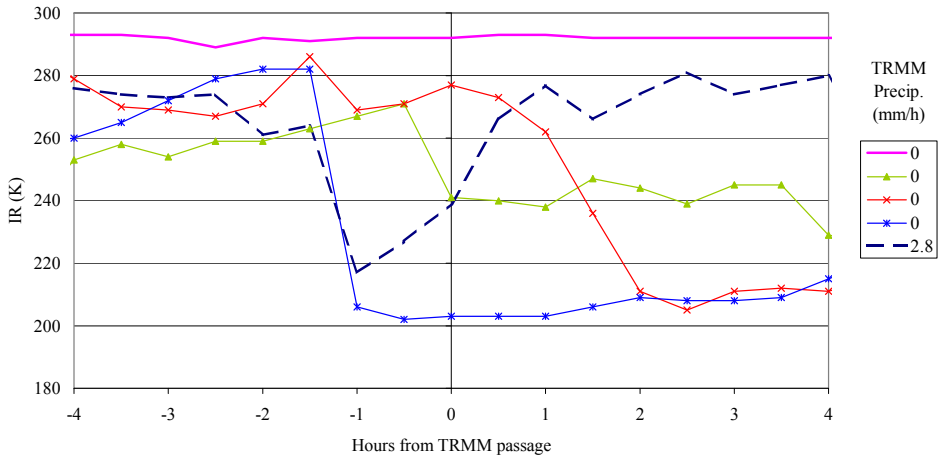


Figure 5.5 IR traces for non convective patterns.

5.2.2 Calibration and Validation of Artificial Neural Networks

The performance of the ANN architectures described in section 5.2.1 were evaluated using the following indicators:

$$AWES = 1 - POD + FAR$$

$$POD = \frac{Cc}{Cc + Nc} \qquad (5.1)$$

$$FAR = \frac{Cn}{Nn + Cn}$$

Where: AWES = Area-Weighted classification Error Score (Todd et al., 1995);

POD = Probability of Detection;

FAR = False Alarm Rate;

Cc = number of convective patterns detected as convective by the MFFNN;

Cn = number of non-convective patterns detected as convective by the MFFNN;

Nc = number of convective patterns detected as non-convective by the MFFNN;

Nn = number of non-convective patterns detected as non-convective by the MFFNN.

The verification set was developed independently of the set used to calibrate the MFFNN, but it also consisted of the same type of convective and non-convective patterns. The calibration set was also used to determine model biases and weights. Thus, to assess model performance, it became necessary to use yet a third set of patterns completely independent from the training and verification sets. The third set will be referred to as the validation set and was developed from all IR patterns coincident with a TRMM PR convective pixel during the last 20 days of each month of 1998. It also included all non-convective patterns associated with non-convective TRMM PR pixels at a distance of three or more pixels from any fully or partially convective pixel during the same period.

In all configurations, verification POD was above 0.98 and FAR below 2%, but shorter input traces featured higher POD and AWES, probably because reduced input lengths are less affected by the variability in storm duration and evolution. However, this trend stopped or even reversed at the 5-value input configuration. The Verification FAR

did not seem to depend significantly on the length of the input. There was no apparent relation between the number of ANN layers and nodes and its performance over the verification set, other than the observation that the two layer MFFNNs with many nodes in the first hidden layer performed somewhat worse than did the rest of the ANNs.

Also impressive wereValidation statistics, given that they were derived from a completely independent set of data: POD is in the 78-81% range, a high value considering that the validation set also included isolated fully convective pixels and percentages of "shallow convection"; "deep, but slowly building convection"; and "Not usable" patterns likely similar to those in the training/verification period. The validation FAR is notably low, ranging around 3-5%. Similarly to verification, validation POD and FAR decreased with the number of input values.

The best results were obtained with 6-value inputs (Figure 5.6) followed by 5-value inputs. This is positive because a reduced input set is less computationally demanding and less affected by missing values. It also reduces the delay in providing real-time precipitation estimates. Among all ANN architectures with 6 input values, the best was the one with two hidden layers of six nodes each. Its validation performance is as follows:

$POD_{valid} = 0.81$; $FAR_{valid} = 0.04$; $AWES_{valid} = 0.23$

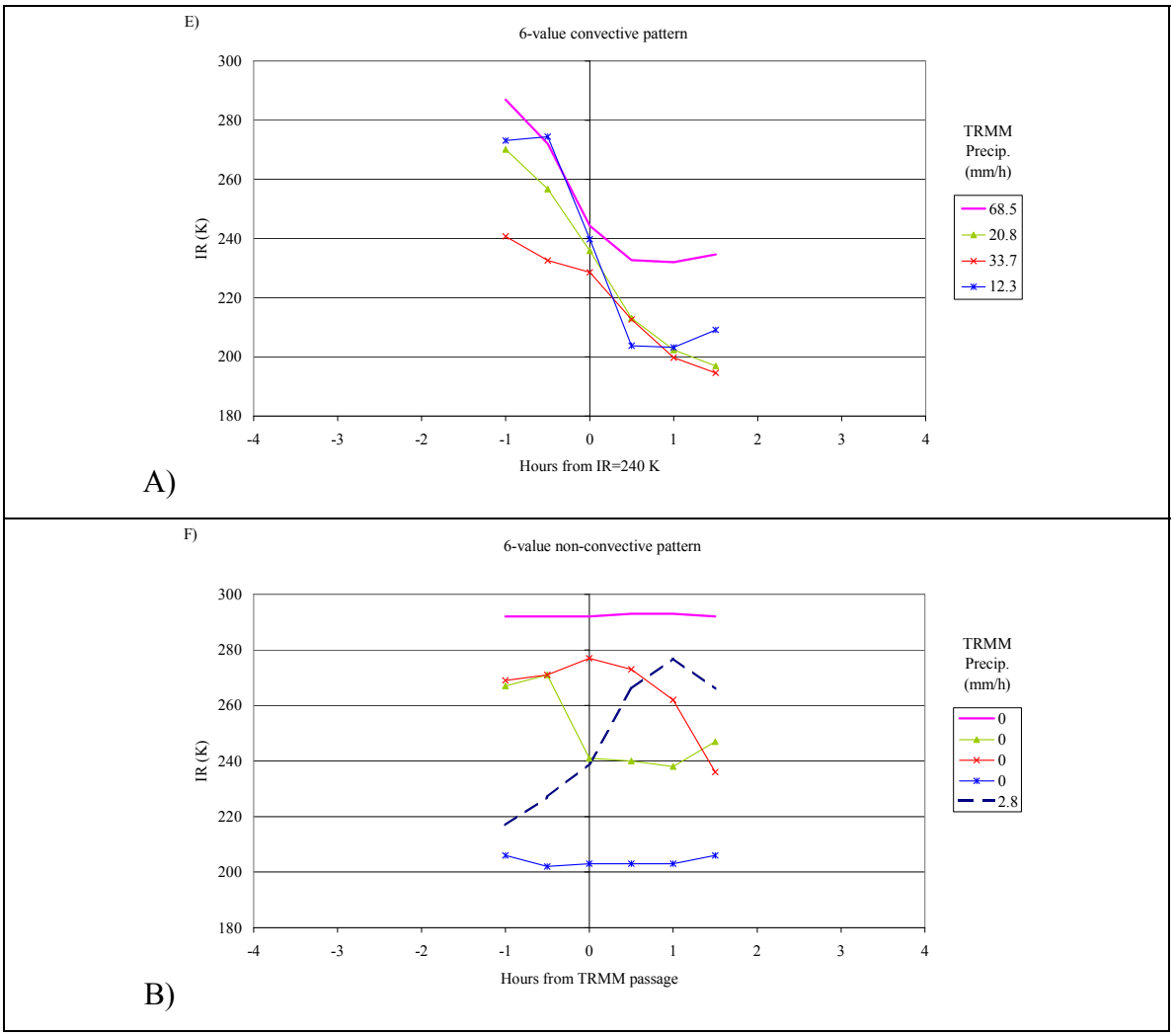


Figure 5.6 Convective (A) and non-convective (B) traces used for training six-value neural networks.

5.3 Identification of the Storm Stages

The frequency of TRMM measurement of precipitation at any given location around the equator varies between once a day for TMI to once every three to four days for PR. Consequently, TRMM images provide a spatial distribution of precipitation, but cannot directly provide its temporal distribution. Geostationary IR signals in pixels belonging to the same storm, on the other hand, when properly synchronized show a strong similarity (Figure 5.7). It can be argued that, at any given time, neighboring pixels are experiencing different temporal phases of the same storm evolution. Therefore, the

TRMM snapshot of precipitation can be also viewed as an image of the rain-rates associated with different temporal stages of the storm. This means that a proper synchronization of the TRMM samples according to the storm stage, rather than according to the measurement time, allows for the reconstruction of the temporal distribution of precipitation for the storm from the TRMM image (Figure 5.7).

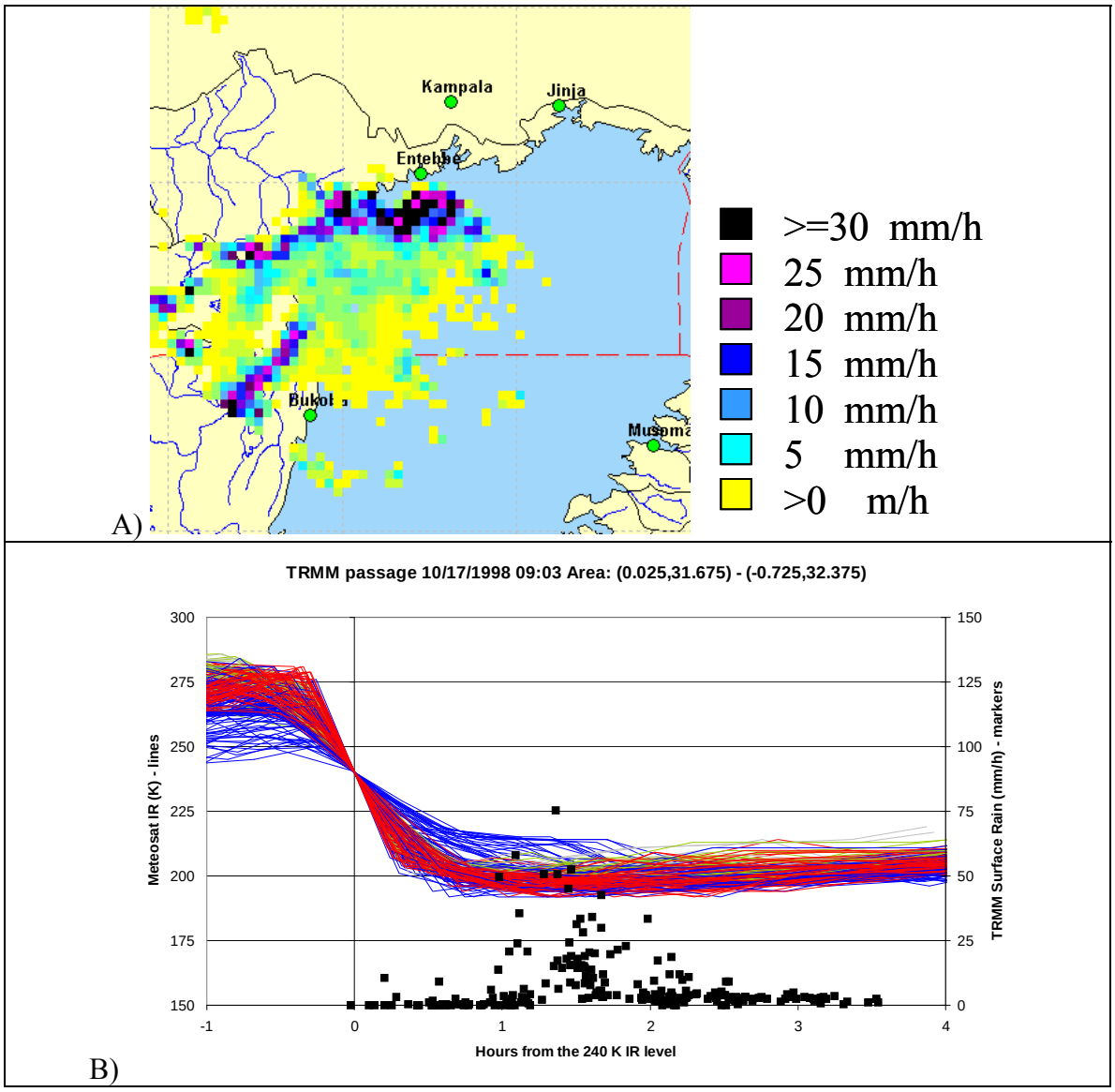


Figure 5.7 A) TRMM PR spatial rain distribution; B) Temporal precipitation distribution derived from synchronizing the corresponding IR patterns around the IR=240 °K point.

A good point for synchronizing the storm IR traces appears to be the time when IR reaches 240 °K during the IR dip, which is characteristic of the developing phase of the convective cell (Figure 5.7). Since the traces used for training the neural network for recognizing the onset of a convective storm were synchronized around this temperature, the time when the neural network peaks can also be used as a synchronizing point. In this case, the neural network not only identifies the presence of a convective storm over a pixel, but also defines its stage evolution index by the number of half-hour timeslots from the neural network peak.

It is assumed that the storm ceases to affect a pixel when the pixel IR temperature returns above 253 °K (-20 °C), a value often used to separate stratiform clouds from non-raining clouds (Griffith, 1987; Adler and Negri, 1988). Since the signals used to train the neural network have a lead-time of one to two time-steps, this rule is enforced only for time-slots that are at least three time-steps from the peak of the neural network. Furthermore, the storm stage index varies from minus one (30 minutes before the Neural Network indicates the convective event) to twenty (ten hours later). The precipitation rate at the twentieth time slot after the neural network peak is used for the rare cases when the storm lasts more than ten hours.

5.4 Summary

A typical IR temporal pattern can be associated with the temporal evolution of ideal convective storms at the pixel level. Real life cloud dynamics confers considerable variation to the IR temporal patterns actually associated with convective precipitation. To account for such a deviation from the ideal model, coincident TRMM PR rain type data and coincident sequences of consecutive IR data from geostationary satellite were used to create a database of IR signals corresponding to convective and non-convective pixels. IR patterns associated with convective pixels were synchronized around a common point close to the onset of the storm in order to generate a family of similar and

easily recognizable IR temporal patterns. An MFFNN was successfully trained to separate convective IR patterns from non-convective ones. The best performances were obtained with sequences of six consecutive IR values beginning one to two time-slots before the dip in the IR signal that characterizes the developing phase of a convective cell. The convective storm activity over a pixel ends when the IR temperature of the pixel becomes warmer than 253 °K. The difference in time from the peak in the MFFNN ‘s output is used to characterize the temporal evolution of the convective storm.

CHAPTER 6

CHARACTERIZATION OF PRECIPITATION RATE DISTRIBUTIONS

6.1 Chapter Scope

The core of indirect and mixed estimation schemes is the relation associating the observed radiation with the precipitation rate. In this research, the rain-rate distributions as a function of IR, VIS/WV, storm stage, orography, and month have been derived from 816 TRMM passages over the Lake Victoria area during 1998-1999. This database consists of more than 1.9 million points. Section 6.2 discusses the influence of storm stage, IR, and orography on the average precipitation rates. Traditional precipitation estimation procedures typically use this type of relations to estimate precipitation. Conversely, probabilistic estimation procedures need to associate an entire precipitation distribution with the observed radiation. Section 6.3 analyzes the merits of using the sample distributions for such a purpose. Section 6.4 examines the advantages of using analytical descriptions of the precipitation distributions, where two analytical models are fit to the observed distributions. Finally, Section 6.5 summarizes the Chapter.

6.2 Relation between IR, Stage, Terrain, and Precipitation

Figure 6.1 and Figure 6.2 show the average precipitation rates derived from 1998-1999 TRMM PR surface rain data as a function of IR and storm stage respectively for Lake Victoria pixels and for land pixels between 2000m and 3000m elevation. The "All Pix" curves include all available pixels without distinction of storm stage. The "Non-Cl" curves represent the precipitation rate of pixels not identified by the neural network as belonging to a convective storm.

Chart A shows that the convective storm classification captures the large majority of intense-rain pixels since the "Non-Cl" precipitation is mostly lower than the average precipitation. The second slot after the neural network peak corresponds to more or less the middle of the IR decrease during the development phase and shows high precipitation intensity over the entire IR range, with increased strength in the lower IR.

The following slots (Chart B) have precipitation curves of similar shape, but intensity decreasing with the distance from the neural network peak. Remarkably, precipitation intensity for the same IR varies considerably with stage number, confirming the assumption that precipitation rate depends on storm stage. In addition, the shapes of the precipitation curves vary considerably with orography. Rain intensity at low IR is higher over the lake than over the land, while the reverse is true for IR above 230 °K.

Figure 6.1 and Figure 6.2 show that the precipitation distributions in the late mature and decaying phases (stages later than NN+4) vary only gradually and reach an almost homogeneous distribution in very late stages. Since the number of samples decreases with the distance from the peak of the neural network due to the extinction of the smaller storms, the distributions for stages later than "NN+4" have been merged into five larger distributions to ensure more accurate statistics, namely:

"NN+5","NN+6", and "NN+7";
"NN+8", "NN+9", and "NN+10";
"NN+11", "NN+12", and "NN+13"
"NN+14", "NN+15", and "NN+16"
"NN+17", "NN+18", "NN+19", and "NN+20".

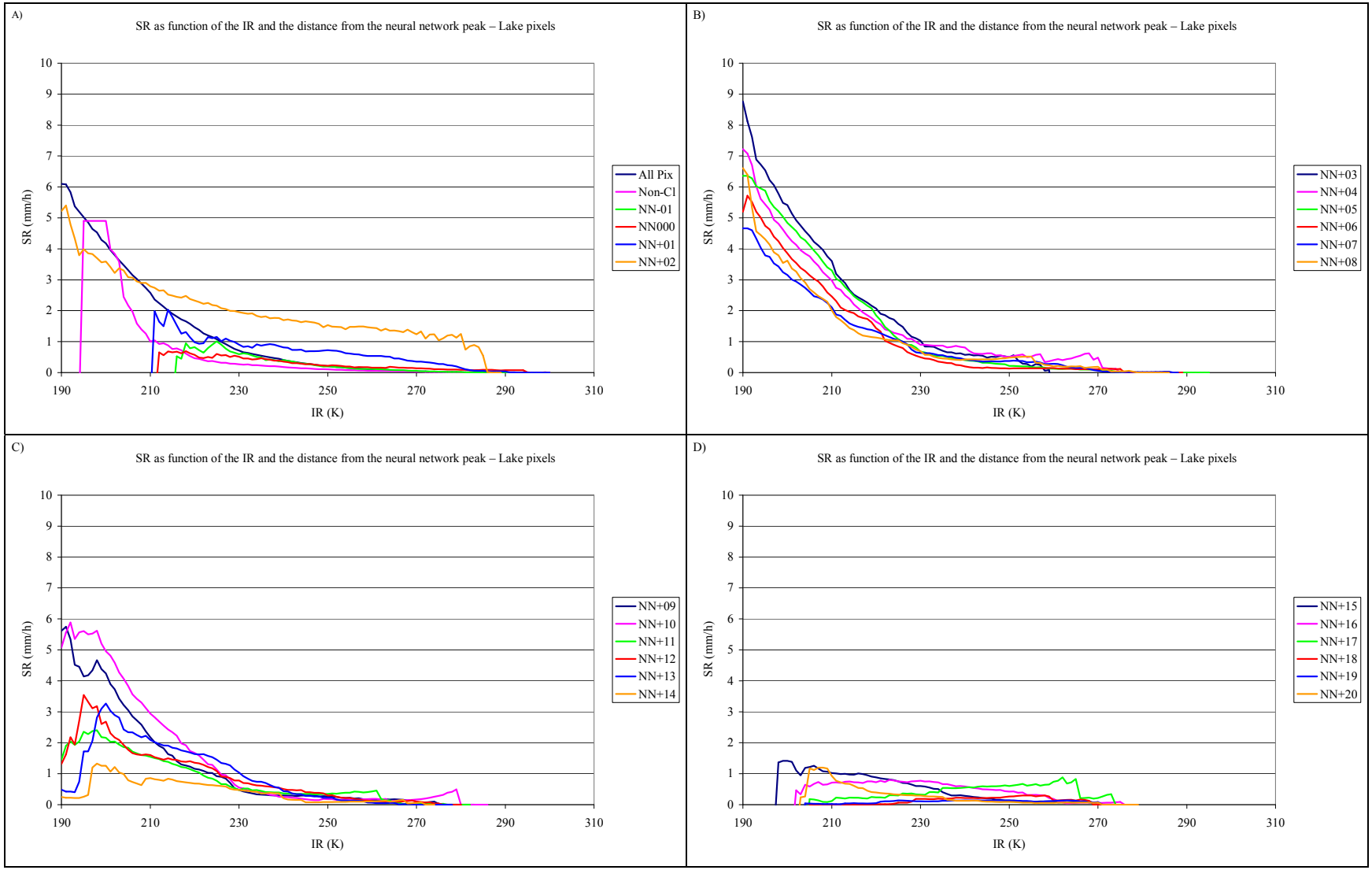


Figure 6.1 TRMM PR surface rain distribution as a function of IR and distance from the neural network peak (stage) for Lake pixels.

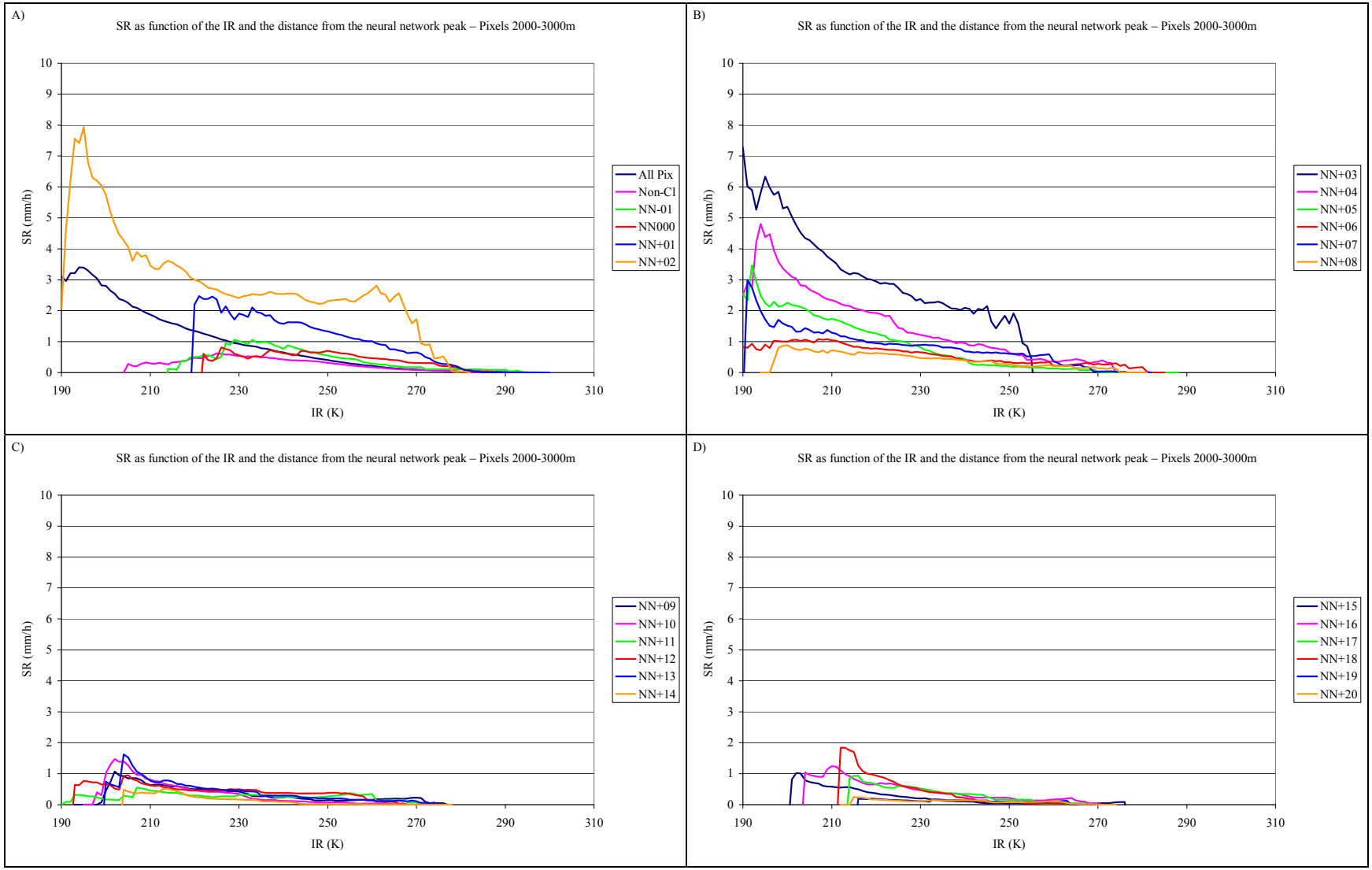


Figure 6.2 TRMM PR surface rain distribution as a function of IR and distance from neural network peak (stage) for 2000-3000 m pixels.

6.3 Empirical Distributions

The distributions reported in Figure 6.1 and Figure 6.2 show the average precipitation rate as a function of terrain, storm stage, and IR temperature. This type of curve could be used as basis for a deterministic precipitation estimation procedure, such as the ones described in Sections 2.3.3.1 and 2.3.3.3. In the probabilistic approach to remote sensing of precipitation, however, ensembles of precipitation rates are obtained at each time step by sampling the precipitation cumulative density function (cdf) associated with the combination of observed input variables. The precipitation cdfs can be described as sample or empirical distributions or as analytical distributions. In the first case, for each combination of the input variables, the sample cumulative distribution is built from the observed data and used as the cdf of the precipitation rate. In the second case, a theoretical probability model is fit to the observed data and used to represent the precipitation cdf in the probabilistic estimation algorithm.

The advantage of the first approach is that it actually describes the observed data, without introducing spurious approximations and extrapolations. Furthermore, if the number of data used to compute the sample cumulative distributions is sufficiently high, the resulting cdf should be sufficiently representative of the real precipitation distribution. Finally, sample cdf are easy to update and are relatively robust since outliers influence only a small number of cases, not the entire distribution.

The most obvious disadvantage of the empirical distributions in the context of the presented methodology, however, is the very large amount of computer memory they require. This is because the empirical distribution corresponding to each combination of the five input variables is made of hundreds of data, and, depending on the desired resolution of the IR/WV/VIS variables, the number of possible combinations may easily approach one million. A second disadvantage is that modeling the temporal and spatial

correlation of data described by empirical distributions is very cumbersome and approximate (Sections 7.4.1 and 8.6.1).

Similarly, the only way to localize the empirical distributions in time and/or space is by building them using only data from the period/area of interest. This, however, means that it becomes necessary to reduce the detail of the observed radiation/ precipitation relation.

In order to save computer memory, two simplifications were made to the way the empirical distributions are treated. First, the TRMM precipitation rates were rounded to the nearest integer value. Second, the maximum admissible precipitation rate was set to 100 mm h^{-1}. With these simplifications, each cdf is represented by 100 values, independently of the number of samples used to compute it. Because only a few dozen of the TRMM samples in the database were larger than 100 mm h^{-1}, mostly by few millimeters, the impact of these simplifications on the precipitation estimation should be limited. In addition, due to the fact that the TRMM PR retrieval algorithm does not consider the increase in reflectivity caused by hailstones, which are sometimes present in tropical convective storms, especially at high elevations, it is possible that at least some of the rain rates above 100 mm h^{-1} were partially overestimated.

Two examples of the precipitations distributions resulting from these simplifications are reported in Figure 6.3. The charts highlight the influence of terrain and IR temperature on the precipitation distribution during the mature phase of the storms. The fact that the shapes of the precipitation distributions over the lake are rougher than the corresponding curves for land pixels is due to the smaller number of available lake pixels.

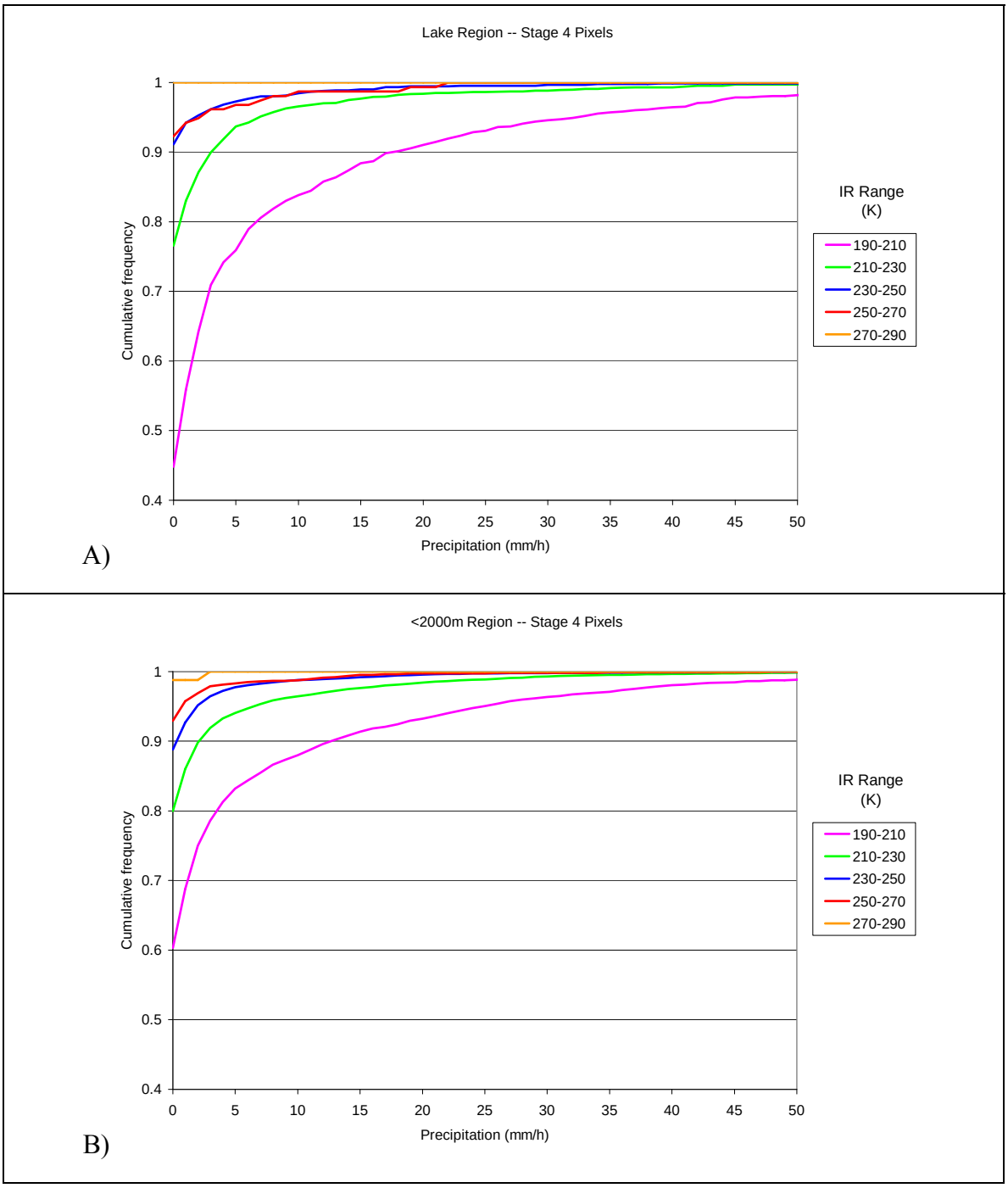


Figure 6.3 Empirical probability distribution as a function of IR temperature for storm stage 4 pixels. A) Lake pixels; B) Land pixels with elevation of less than 2000 m.

6.4 Analytical Distributions

Analytical formulations of the precipitation probability distribution solve most of the disadvantages of the empirical distributions:

1. Only a handful of parameters must be preserved for each combination of the input variables, decreasing the amount of data that must be kept in the computer memory.
2. Spatial and temporal correlation is easily implemented operating on the parameters of the distribution, usually the mean and variance.
3. Bayesian approaches similar to equations 2.15 - 2.16 can be used to create local versions of the distributions without loss of detail in the observed radiation/precipitation relation.

However, analytical distributions also have their disadvantages, especially in the framework used in this research:

1. It may be difficult to find a single mathematical model that adapts to the variety of distributions considered in this research.
2. The difference between sample and analytical cdfs may be relevant, possibly causing a decrease in estimation performances. This is especially true at the upper limit of the observed cdf, where the analytical model may not be able to comply with the physical limits imposed on the precipitation rates by the observed cloud configuration.
3. Outliers may influence the computation of the parameters of the analytical model and affect the entire analytical distribution. This is especially important when the parameters are computed with a limited number of values.

Two analytical models were adapted to the estimation framework used in this research, namely, that proposed by Bell (1987) and that used by Fiorucci et al. (2001).

Bell's model considers the precipitation probability density function as composed of two components, a discrete probability of no-rain and a lognormally distributed probability of positive rain.

$$f_R(z) = \begin{cases} 0 & z < 0 \\ P(0) & z = 0 \\ P(0) + N(\ln(z), \mu_{LNR}, \sigma_{LNR}) * (1 - P(0)) & z > 0 \end{cases} \qquad (6.1)$$

Where: $f_R(z)$ = probability density function of precipitation R;

$P(0)$ = probability of no rain;

$N(\ln(z), \mu_{LNR}, \sigma_{LNR})$ = lognormal distribution of positive rain rates;

In the alternative model, the discrete probability of no-rain is not considered. Therefore, the precipitation is described as lognormally distributed (Fiorucci et al., 2001). Furthermore, the variance is set equal to the mean squared.

Figure 6.4 shows that the Bell's model fits the sample distribution very well, both for convective conditions, dominated by intense precipitation, and for stratiform conditions, dominated by mild and negligible precipitation. However, when intense precipitation is frequent, Bell's model seems to overestimate the probability of intense rain. The simpler model by Fiorucci et al. (2001) is not nearly as accurate, overestimating the probability of low rain and underestimating the probability of intense rain.

Overall, Bell's model seems able to reproduce the sample distributions in a variety of situations, at least when the number of samples used to compute the mean and variance of positive rain rates is relatively large. One possible problem with this model is

that two of its three parameters are computed using only positive precipitation rates, which for some combinations of the input variables are just a very small fraction of the TRMM data.

6.5 Summary

Distributions of precipitation rates as a function of the IR, VIS/WV, storm stage, month, and terrain are of fundamental importance in realizing a probabilistic procedure for remote sensing of precipitation. This chapter illustrated the influence of input variables on precipitation distributions. It also examined the pros and cons of using empirical precipitation distributions versus using analytical models. The model by Bell (1987) was able to satisfactorily fit a range of precipitation distributions corresponding to different combinations of the input variables. Analytical models are more flexible and allow an easier implementation of spatial and temporal correlations. However, the fit of the sample distributions is not always perfect even in presence of a large number of samples. For this reason, the use of the empirical distributions cannot be discarded a priori.

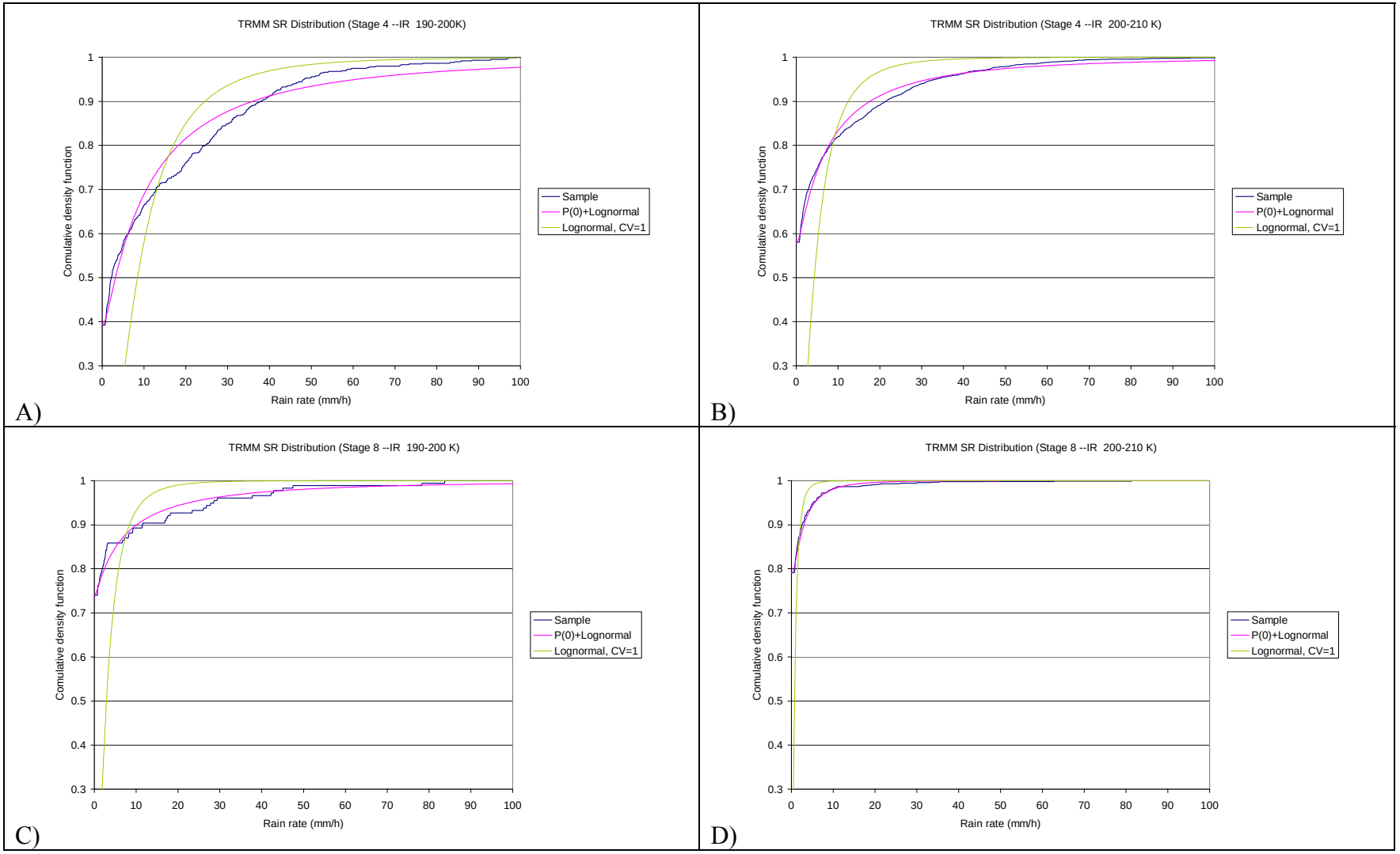


Figure 6.4 Empirical and analytical precipitation distributions for convective precipitation (A, B) and stratiform precipitation (C, D).

CHAPTER 7

SINGLE-PIXEL ANALYSIS

7.1 Chapter Scope

Precipitation estimation at the pixel level is obtained by integrating the identification and temporal evolution of convective storms described in Chapter 5 within the estimation structure described in Chapter 3. Single-pixel analysis is important because it allows a quick calibration of the estimation procedure and its validation over a large number of gages, covering different precipitation patterns. This chapter is organized as follows: Section 7.2 describes the statistics used to assess the model's ability to represent precipitation variability over single pixels in the basin. Section 7.3 begins by examining some implementation details of the estimation process. Then, it investigates the calibration, validation, and sensitivity of the precipitation estimation using the precipitation empirical distributions and using the analytical formulation of the precipitation distributions. Section 7.4 addresses the problem of incorporating the precipitation temporal autocorrelation in the estimation process. This is done first with regard to the empirical precipitation distributions and then for the analytical formulation. Estimation performances of the temporally correlated models are further discussed in Section 7.5. Finally, Section 7.6 summarizes the research findings and discusses possible extensions.

7.2 Performance Indicators

To assess the effectiveness of the proposed precipitation estimation approach, two aspects must be considered at the same time:

- The ability of the ensemble average to track the behavior of the gage records

- The capability of the ensemble to represent in a statistically meaningful way the uncertainty of the precipitation patterns

The first aspect is quantified by commonly used statistics:

Bias (Bias): For a single gage this is the difference over the simulation period between the total gage precipitation and the average of the ensemble total precipitation. This value is given as a percentage of the total gage precipitation given the large variation in the precipitation patterns around the basin. When applied to a group of gages, this statistic is the average of the single rain gage relative biases.

Absolute Bias (Abias): This term is used only when considering a group of rain gages. It is the average of the absolute value of the single rain gage relative biases. It gives a measure of the spread of the biases.

Correlation (Corr): For a single gage, this is the correlation coefficient between the ensemble average and the corresponding rain gage record over the same estimation period (day, dekad, or month). When applied to a group of gages, this statistic is the average of single rain gage correlation.

Rank correlation (Rcorr): For a single gage, this is the correlation coefficient between the rank of the ensemble average and the rank of the corresponding rain gage record over the same estimation period (day, dekad, or month). When applied to a group of gages, this statistic is the average of single rain gage rank correlation.

Mean Absolute Error (MAE): For a single gage, this is the average of the absolute difference between the gage record and the ensemble average over the same estimation

period (day, dekad, or month). This value is given as a percentage of the average gage precipitation. When applied to a group of gages, this statistic is the average of the single rain gage MAE.

The capability of the ensemble to represent in a statistically meaningful way the precipitation variability at the gage is quantified by the following statistics:

Average ensemble 95% width (95%R): For a single gage, this is the average of the differences between the 2.5-percentile and the 97.5-percentile of the estimate ensemble. When applied to a group of gages, it is the average of the single rain gage 95% widths.

Ensemble 95% compliance (95%C): For a single gage, this is the frequency at which the gage records fall between the 2.5-percentile and the 97.5-percentile of the ensemble. When applied to a group of gages, it is the average of the single rain gage 95% compliances.

Lower end 95% error (95%L): For a single gage, this is the frequency at which the gage records are less than the ensemble 2.5-percentile. When applied to a group of gages, it is the average of the single rain gage lower end 95% error.

KS-statistics (KS): For a single gage, this is the maximum of the absolute difference between the distribution of the ensemble-percentiles corresponding to the gage readings and the 0-1 uniform distribution. The rationale behind this index is that the ensemble represents the variability of precipitation in a statistically meaningful way if the gage position with respect to the ensemble is uniformly distributed over the entire range. When applied to a region, it is the average of the single rain gage KS statistics.

90% KS acceptance (KS10): For groups of rain gages, it is the number of gages that have a KS statistic below which rejecting the hyporesearch that the percentile distribution is 0-1 uniform implies a 10% error probability.

99% KS acceptance (KS01): For groups of rain gages, it is the number of gages that have a KS statistic below which rejecting the hyporesearch that the percentile distribution is 0-1 uniform implies a 1% error probability.

7.3 Characteristics of the Single Pixel Estimation Process

7.3.1 Characteristics of the Estimation Algorithm

Empirical distributions and analytical distributions

The empirical distribution model (EIVWNUU) was implemented first and used for calibration and validation. The procedure adopted to build the sample distributions of precipitation rates is the following:

1. All available TRMM PR data are partitioned into subsets according to the month, with the constraints described later;
2. The subsets of TRMM data obtained in step 1 are further subdivided into subsets according to the terrain classification of Table 7.1;
3. The subsets of TRMM data obtained in step 2 are partitioned into subsets according to the storm stage classification described in Section 6.2.
4. Daytime pixels belonging to the subsets obtained in step 3 are subdivided into subsets according to their IR and VIS. Each subset contains samples belonging to rectangular regions of the IR/VIS domain. The size of the IR/VIS regions is chosen so that each region contains a number of elements that yields a reliable cdf. Further, the regions must form a partition of the IR/VIS domain (Figure 7.1).

5. Similarly, nighttime pixels belonging to the subsets obtained in step 3 are subdivided into subsets according to their IR and IR-WV. Each subset contains samples belonging to rectangular regions of the IR/IR-WV domain. The size of the IR/IR-WV regions is chosen so that each region contains a number of elements that yields a reliable cdf. Further, the regions must form a partition of the IR/IR-WV domain. Since WV data were frequently missing, a second family of subsets was created, partitioning the nighttime pixels according to their IR value alone.

This procedure is necessary because TRMM data are not uniformly distributed in the storm stage/IR domain, as shown in Figure 6.1 and Figure 6.2, and, consequently, in the storm stage/IR/VIS and storm stage/IR/IR-WV spaces. Thus, a regular partitioning of these two spaces would have produced either cdf computed with an insufficient number of elements or a loss of resolution in the IR, VIS, and IR-WV dependencies.

The minimum number of data to be used for computing a sample cdf is found by maximizing the satellite-to-gage correlation. After extensive experimentation, the minimum number of elements that must be used to create a sample-cdf was set to 75. In most cases, the actual number of data used to create the sample cdf is above this value.

The analytical model (AIVWNUU) was implemented in a later phase using the same settings used for the EIVWNUU procedures, but fitting the analytical model to the sample distributions. However, the minimum number of elements used to determine the distribution parameters was set to 150 to increase the reliability of the distribution parameters.

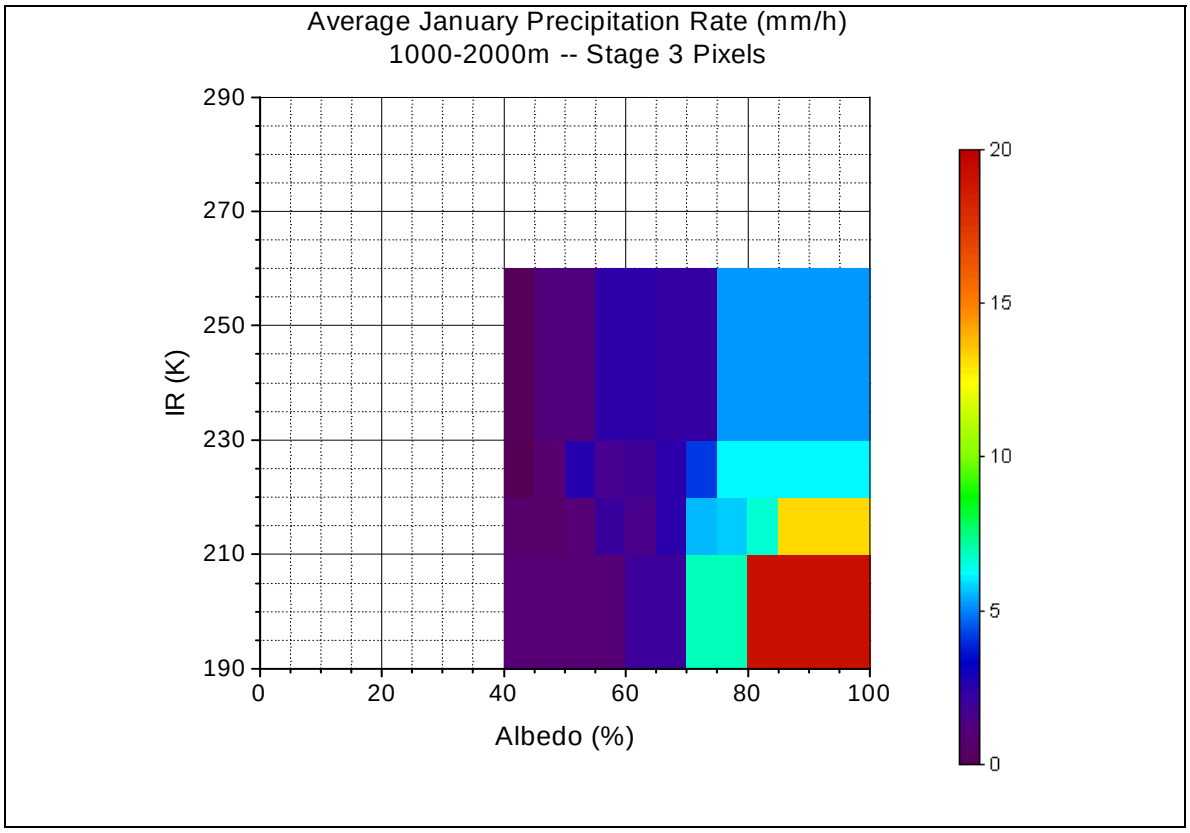


Figure 7.1 Example of partition of the IR/VIS domain to produce the sample distributions.

Orography

Sections 6.2 and 6.3 showed that there are deep differences in the precipitation distributions over different geographical areas. Incorporating these differences in the estimation procedure translates into better precipitation estimates. Among several different partitions of the TRMM data according to orography, the one reported in Table 7.1 has proved to be most robust.

Table 7.1 Partitioning scheme of the TRMM pixels according to the pixel orography.

Region of application	Lake	0-1000m		1000-2000m		2000-3000m	
TRMM SR pixels for precipitation distribution		min	max	min	max	min	max
	Lake	0	1250	750	2250	1750	3250

This partitioning scheme subdivides the testing area into four regions: the Lake Victoria pixels, the land pixels below 1000 m, the pixels between 1000 and 2000 m, and the pixels between 2000 and 3000 m. Since there are no rain gages above 3000 m, this area is not considered here. In an operational use of the procedure, a fifth region considering these pixels must be added.

Only SR data from the Lake Victoria pixels are used for deriving the rain-rate distribution used over the lake region. The rain-rate distributions for the three elevation bands of the land pixels are derived using SR data from pixels up to 250 m above and below the elevation band boundaries. This extension is necessary to include enough data to yield sufficiently detailed precipitation distributions even when several input variables are considered. The amount of TRMM data in 1998-1999 is not sufficient to design a more detailed partitioning scheme.

Monthly distributions

There are several reasons why it is useful to include a certain degree of seasonality in the choice of the rain-rate distribution:

- In certain periods of the year cold, but non-rainy, clouds like the cirri may be more common than the cold, but rainy, clouds like the cumuli or nimbo-cumuli;
- Availability of humidity in the air may affect the precipitation generated from similar types of clouds. This is especially true during the monsoon and the dry seasons.
- Albedo has a seasonal pattern that is only partially compensated by the solar angle correction.

The approach taken in this research is to generate a different rain-rate distribution for each month. However, given the limited amount of data available, using a truly

monthly distribution causes a decrease in the detail of the relation between rain-rate and IR/VIS or IR/WV, resulting in better statistics at the monthly level, but worse at the daily. As a compromise, the "monthly" precipitation distributions are therefore generated using TRMM data from the specific month in question and from the month before and after. Further, the annual distribution is used in place of a "monthly" distribution if the latter does not contain enough data to achieve at least some basic partition of the Storm stage/IR/VIS and Storm stage/IR/WV domains.

IR and VIS screening

Meteosat images are affected by thermal noise and sometimes by slight changes in the satellite position and inclination, which can cause geometric error in the images. Further, several Meteosat images were missing and were replaced by data obtained from linearly interpolating the closer available images. TRMM PR surface rain data are also affected by thermal noise and may be affected by spurious echoes, especially over mountains. The resampling of the Meteosat and TRMM images to a regular grid introduces a further source of error. Finally, and likely most importantly, TRMM and Meteosat images are not always coincident, but may be spaced up to 15 minutes apart. All these errors cause precipitation rates to be wrongly attributed to high IR and low VIS, thus increasing the estimation bias and decreasing its correlation. To decrease the impact of this phenomenon, precipitation rates corresponding to IR > 258 ºK and VIS < 40% have been screened out. The IR screening was set higher than the threshold used to determine the end of the convective activity (Section 5.3) in order to accommodate shallow convection and orographic rain.

Number of traces in the precipitation ensemble

Increasing the size of the ensemble reduces the uncertainty in the ensemble statistics introduced by the random selection of the half-hour precipitation rates which make up the ensemble. By contrast, increasing the size of the ensemble also linearly increases the computation time. Table 7.2 shows that a minimum of 200 ensemble

elements is required to achieve stable simulation results over long temporal horizons. However, the size of the ensemble has been set to 500 elements to maintain a safe margin.

Table 7.2 Variation of precipitation estimation statistics using the median of the ensemble for different ensemble sizes.

	Entire period	Daily			Dekad			Month		
Ensemble elements	Bias/ Gage	Corr	MAE/ Gage	95%C	Corr	MAE/ Gage	95%C	Corr	MAE/ Gage	95%C
100	0.043	0.444	1.097	0.910	0.647	0.552	0.806	0.744	0.385	0.763
200	0.043	0.446	1.095	0.903	0.648	0.552	0.792	0.743	0.386	0.734
500	0.043	0.447	1.096	0.901	0.648	0.552	0.799	0.743	0.386	0.737
1000	0.044	0.447	1.095	0.906	0.647	0.552	0.794	0.743	0.386	0.739

7.3.2 Model Calibration (1996-1998)

The calibration dataset is composed of 60 gages distributed over the elevation range 630-2300 m and located mostly in the northern and western sides of the basin (Figure 4.2 A). It should be noted that some parameters of the estimation procedure described earlier (i.e., IR and VIS thresholds, number and limits of the orographic bands, use of tri-monthly precipitation distribution, among others) were selected to optimize the results from 1996 to 1998. In selecting these parameters, however, care was exercised to ensure that this selection was not too specific to the calibration dataset. For example, better results could have been obtained if the elevation bands used for regionalizing the rain-rate distributions were 0-1100, 1100-2000, and 2000-3000 meters instead of the more regular 0-1000, 1000-2000, 2000-3000 m actually adopted. Furthermore, slightly better results would have been obtained if VIS screening had been performed at 41%

instead of 40%. However, such improvements would have been minor and would not have added to the general applicability of the method.

Table 7.3 reports the performance of the single-pixel estimation procedure (EIVWNUU) compared against the performance of the single-pixel adaptation of the GPI technique. According to the latter, the rain-rate equals 3 mm h^{-1} if IR is below 235 °K, and zero otherwise. Results show that the proposed model has a much better bias than the GPI over the period 1996-1998 and also much better correlation, especially at the daily level.

Table 7.3 EIVWNUU and GPI performance for the calibration dataset (1996-1998).

Day Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
GPI	0.75	0.75	0.38	0.51	1.57						
EIVWNUU	0.04	0.13	0.45	0.57	1.10	14.9	0.91	0.02	0.36	0	0
Dekad Method	**Bias/ Gage**	**Abias /Gage**	**Corr**	**RCorr**	**MAE/ Gage**	**95%R**	**95%C**	**95%L**	**KS**	**KS_{10}**	**KS_{01}**
GPI	0.75	0.75	0.60	0.68	0.96						
EIVWNUU	0.04	0.13	0.65	0.72	0.55	64.8	0.80	0.12	0.21	27	40
Month Method	**Bias/ Gage**	**Abias /Gage**	**Corr**	**RCorr**	**MAE/ Gage**	**95%R**	**95%C**	**95%L**	**KS**	**KS_{10}**	**KS_{01}**
GPI	0.75	0.75	0.71	0.73	0.84						
EIVWNUU	0.04	0.13	0.74	0.75	0.39	121.7	0.74	0.16	0.24	22	42

However, even though the 95% compliance rate might be acceptable at the daily level, it is low at the dekad and monthly level.

7.3.3 Model Validation (1996-1997)

The validation dataset is composed of 42 gages, which were not included in the calibration dataset. These gages are distributed over the elevation range 816-2400 m and

are located mostly in the southern and eastern sides of the basin, with three gages situated on islands (Figure 4.2 A and B).

Table 7.4 shows that the proposed model (EIVWNUU) has a bias slightly higher than the bias of the calibration set, but clearly lower than the GPI bias. The correlation is also higher, reaching almost 0.80 at the monthly level, and continues to be better than GPI's correlation, especially at the dekad level. The 95% compliance rate for the validation case is a little lower than it is for the calibration case. Overall, the validation statistics are very similar to the calibration statistics, supporting the hyporesearch that the estimation procedure is generally valid, independently of the calibration data set.

Table 7.4 EIVWNUU and GPI performance for the validation dataset (1996-1997).

Day Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
GPI	0.87	0.89	0.41	0.50	1.69						
EIVWNUU	0.08	0.19	0.47	0.55	1.14	13.3	0.91	0.03	0.43	0	0
Dekad Method	**Bias/ Gage**	**Abias /Gage**	**Corr**	**RCorr**	**MAE/ Gage**	**95%R**	**95%C**	**95%L**	**KS**	**KS_{10}**	**KS_{01}**
GPI	0.87	0.89	0.63	0.72	1.10						
EIVWNUU	0.08	0.19	0.70	0.76	0.60	60.1	0.77	0.15	0.25	17	28
Month Method	**Bias/ Gage**	**Abias /Gage**	**Corr**	**RCorr**	**MAE/ Gage**	**95%R**	**95%C**	**95%L**	**KS**	**KS_{10}**	**KS_{01}**
GPI	0.87	0.89	0.75	0.77	0.98						
EIVWNUU	0.08	0.19	0.79	0.81	0.43	112.3	0.69	0.21	0.27	19	32

7.3.4 Comparison with Adjusted GPI (TRMM product 3B42).

The GPI method uses a fixed IR threshold of 235 ºK to separate rainy pixels from non-rainy pixels and applies a fixed precipitation rate of 3 mm hr^{-1}. The Adjusted GPI

(TRMM product 3B42) also uses a fixed IR threshold of 235 ºK to discriminate rainy pixels from non-rainy ones, but the rain-rate is calibrated by month and by 1ºx1º region. As illustrated in Section 2.3.3.3, the optimal rain-rate is determined so that the method applied to coincident IR and TRMM TMI images over each area reproduces the total observed monthly TMI precipitation. In Table 7.5 the Adjusted GPI has been derived by applying the appropriate adjusted rain-rate and 235 ºK threshold to the single-pixel IR sequence. Unfortunately, this product has been available only since January 1998, so it cannot be used as comparison for the 1996-1997 period..

Table 7.5 EIVWNUU, GPI, and TRMM 3B42 performance (1998).

Day

Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
GPI	0.67	0.67	0.42	0.53	1.51						
3B42	0.48	0.48	0.41	0.52	1.41						
EIVWNUU	0.05	0.17	0.46	0.58	1.11	13.8	0.91	0.02	0.40	0	0

Dekad

Method	Bias/ Gage	Bias/ Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
GPI	0.67	0.67	0.65	0.70	0.89						
3B42	0.48	0.48	0.63	0.68	0.79						
EIVWNUU	0.05	0.17	0.65	0.71	0.57	61.5	0.80	0.11	0.30	22	39

Month

Method	Bias/ Gage	Bias/ Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
GPI	0.67	0.67	0.75	0.75	0.77						
3B42	0.48	0.48	0.72	0.70	0.64						
EIVWNUU	0.05	0.17	0.74	0.74	0.41	115.9	0.74	0.15	0.27	32	44

The 1998 dataset is composed of 48 rain gages with at least 10 months of valid precipitation data. All of these gages also belong to the calibration data set.

The 3B42 product bias is substantially lower than the GPI bias, but is still much higher than the EIVWNUU bias. In terms of correlation, EIVWNUU performs better than the other two estimators at the daily level, but is slightly worse than the GPI at the dekad and monthly level. It should be noted, however, that monthly correlation with each gage has been computed here with only 10-12 samples.

The fact that the EIVWNUU performance over a fraction of the calibration period is very similar to the performance over the entire calibration period indicates that the model performs consistently.

7.3.5 Parameter Sensitivity

The precipitation estimation procedure examined in this research estimates precipitation by using IR, VIS, WV, month, and terrain as input variables. In this section, only the influence of the last four variables on model performance is examined. Since the IR radiation is fundamental for the precipitation estimate, the sensitivity to this variable is not evaluated. However, the sensitivity to the IR/VIS filtering thresholds is assessed.

7.3.5.1 Influence of Storm Stage, Visible, and Water Vapor

The inclusion of the storm stages decreases the satellite-to-gage bias by 50%, while increasing the correlation by a little more than half percentage point and the MAE by one to two percentage points (Table 7.6). The first effect is due to the fact that the intense precipitation associated with cold IR during the shorter convective phases is not extended to the equally cold stratiform rain. This is also the explanation for the MAE reduction. The limited effect on the correlation may be partly due to the averaging affecting daily or longer precipitation measures. Further, the more detailed relations between IR/VIS/WV and rain rates that can be derived when ignoring the storm stages may compensate for the loss of this information. A larger TRMM database might possibly have revealed a stronger effect of storm have stages on the satellite-to-gage

correlation, because the detail of the IR/VIS/WV relation with precipitation would have not changed. Overall, the influence of the storm stage is similar to that of the inclusion of the WV variable, although this has a much lower impact on the bias.

Table 7.6 EIVWNUU sensitivity to storm stage, WV, and VIS. Calibration dataset (1996-1998).

Day Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.04	0.13	0.45	0.57	1.10	14.9	0.91	0.02	0.36	0	0
No Stage	0.08	0.15	0.45	0.57	1.11	15.1	0.91	0.03	0.35	0	0
No WV	0.05	0.14	0.44	0.56	1.11	15.0	0.91	0.03	0.35	0	0
No WV/VIS	0.14	0.20	0.42	0.53	1.18	16.7	0.91	0.03	0.33	0	0
Dekad											
Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.04	0.13	0.65	0.72	0.55	64.8	0.80	0.12	0.21	27	40
No Stage	0.08	0.15	0.64	0.73	0.57	65.8	0.79	0.13	0.21	27	44
No WV	0.05	0.14	0.64	0.71	0.56	65.0	0.79	0.13	0.20	31	42
No WV/VIS	0.14	0.20	0.61	0.69	0.61	70.2	0.77	0.16	0.22	27	35
Month											
Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.04	0.13	0.74	0.75	0.39	121.7	0.74	0.16	0.24	22	42
No Stage	0.08	0.15	0.74	0.75	0.40	124.4	0.73	0.16	0.26	18	37
No WV	0.05	0.14	0.73	0.74	0.40	122.3	0.73	0.16	0.25	22	40
No WV/VIS	0.14	0.20	0.71	0.72	0.44	131.7	0.71	0.21	0.29	13	29

Because it is very rare to design a procedure that uses only IR and WV data, the influence of the VIS data has been computed together with that of the WV data. These two additional sources of information have a strong effect on estimation performances, reducing the bias by more than two thirds and increasing the satellite-to-gage correlation by three to four percentage points. The increase in correlation at the daily resolution is especially important given its low level. These three variables also affect the ability of

the model to represent precipitation variability as seen by the decrease in the 95% compliance rate.

7.3.5.2 Influence of Terrain

Terrain data halve the overall bias of the estimation and reduces MAE. The effect on the average satellite-to-gage correlation is very low (Table 7.7). Figure 7.2 shows, in greater detail, that ignoring terrain data results in decreased precipitation estimates both over the lake and at elevation higher than 2000 m, but in increased estimates at lower elevations. Lake precipitation estimates decrease because the high precipitation rates associated with the lower IR typical of the lake storms are reduced when mixed with data from the watershed. Similarly, at high elevations the precipitation rates of the frequent warm orographic precipitation decrease when mixed with data from lower elevations.

Table 7.7 EIVWNUU sensitivity to terrain data. Calibration dataset (1996-1998).

Day

Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.04	0.13	0.45	0.57	1.10	14.9	0.91	0.02	0.36	0	0
No Terrain	0.08	0.15	0.45	0.57	1.11	15.2	0.90	0.03	0.36	0	0

Dekad

Method	Bias/ Gage	ABias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS10	KS01
EIVWNUU	0.04	0.13	0.65	0.72	0.55	64.8	0.80	0.12	0.21	27	40
No Terrain	0.08	0.15	0.64	0.72	0.57	65.9	0.79	0.13	0.22	28	40

Month

Method	Bias/ Gage	ABias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS10	KS01
EIVWNUU	0.04	0.13	0.74	0.75	0.39	121.7	0.74	0.16	0.24	22	42
No Terrain	0.08	0.15	0.74	0.75	0.40	124.3	0.72	0.17	0.26	18	38

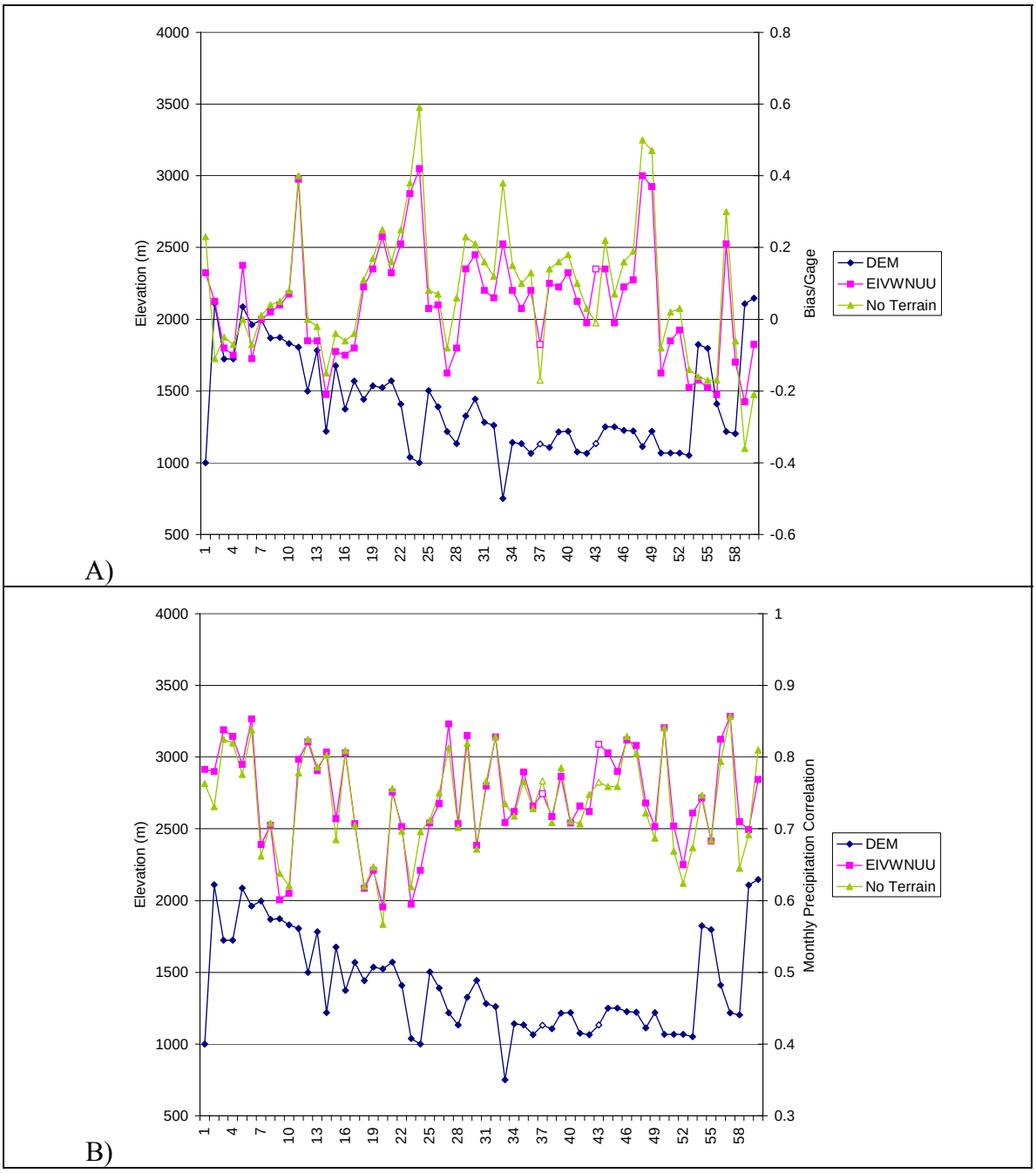


Figure 7.2 Sensitivity of precipitation estimates to terrain data. A) Bias/Gage; B) Satellite-to-Gage correlation of monthly precipitation. Lake gage marks have white interior.

7.3.5.3 Influence of Seasonality

EIVWNUU uses "monthly" precipitation distributions, each derived using three months of data. The use of the same rain distribution for the whole year improves daily estimates because the relations between IR/VIS/storm-stage and precipitation are more detailed, as shown in Table 7.8 by the slight decrease in MAE. On the other hand, the use of annual precipitation distributions significantly worsens satellite-to-gage correlation and the 95% compliance rate over longer periods because differences in cloud dynamics during the dry and wet seasons are not accounted for.

Table 7.8 EIVWNUU sensitivity to exclusion of season data (Annual) and increase in season detail (Bimonthly). Calibration dataset (1996-1998).

Day Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.04	0.13	0.45	0.57	1.10	14.9	0.91	0.02	0.36	0	0
Annual	0.04	0.12	0.45	0.57	1.09	14.7	0.91	0.02	0.36	0	0
Bimonthly	0.05	0.13	0.44	0.56	1.10	15.1	0.91	0.03	0.36	0	0
Dekad Method	**Bias/ Gage**	**Abias /Gage**	**Corr**	**RCorr**	**MAE/ Gage**	**95%R**	**95%C**	**95%L**	**KS**	**KS_{10}**	**KS_{01}**
EIVWNUU	0.04	0.13	0.65	0.72	0.55	64.8	0.80	0.12	0.21	27	40
Annual	0.04	0.12	0.64	0.72	0.56	63.8	0.79	0.13	0.21	29	43
Bimonthly	0.05	0.13	0.65	0.72	0.55	65.4	0.81	0.12	0.21	28	42
Month Method	**Bias/ Gage**	**Abias /Gage**	**Corr**	**RCorr**	**MAE/ Gage**	**95%R**	**95%C**	**95%L**	**KS**	**KS_{10}**	**KS_{01}**
EIVWNUU	0.04	0.13	0.74	0.75	0.39	121.7	0.74	0.16	0.24	22	42
Annual	0.04	0.12	0.72	0.74	0.40	119.6	0.72	0.16	0.24	25	41
Bimonthly	0.05	0.13	0.74	0.75	0.39	123.0	0.74	0.15	0.24	23	42

Conversely, the use of only two months instead of three for computing the "monthly" rain rate distributions decreases estimation accuracy at the daily resolution, while improving it at longer resolutions ("Bimonthly" entry in Table 7.8). This effect is

not very clear because the estimation algorithm uses the annual distributions for the months that do not have enough data to produce detailed relations between precipitation rates and IR/VIS/storm stage data.

7.3.5.4 Influence of IR/VIS Screening

As mentioned earlier, precipitation pixels coincident with IR temperature higher than 258 ºK (-15 ºC) or albedo lower than 0.4 are screened out to decrease the noise introduced by false TRMM radar echoes, temporal and spatial mismatches between TRMM and Meteosat images, antenna noise, and any spatial/temporal interpolation of missing data.

Table 7.9 EIVWNUU sensitivity to IR screening. Calibration dataset (1996-1998).

Day Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.04	0.13	0.45	0.57	1.10	14.9	0.91	0.02	0.36	0	0
Thr_{IR}=253K	-0.02	0.12	0.44	0.56	1.07	14.3	0.90	0.02	0.40	0	0
Thr_{IR}=263K	0.10	0.15	0.45	0.57	1.12	15.6	0.91	0.03	0.32	0	0
Dekad											
Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.04	0.13	0.65	0.72	0.55	64.8	0.80	0.12	0.21	27	40
Thr_{IR}=253K	-0.02	0.12	0.64	0.72	0.54	62.9	0.80	0.10	0.21	28	40
Thr_{IR}=263K	0.10	0.15	0.65	0.72	0.57	66.8	0.79	0.15	0.22	28	40
Month											
Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.04	0.13	0.74	0.75	0.39	121.7	0.74	0.16	0.24	22	42
Thr_{IR}=253K	-0.02	0.12	0.74	0.75	0.38	118.8	0.75	0.12	0.23	27	41
Thr_{IR}=263K	0.10	0.15	0.74	0.75	0.40	124.7	0.72	0.19	0.27	16	34

Table 7.9 shows that an increase of the IR screening to 263 ºK (-10 ºC) would significantly increase bias and MAE, without affecting satellite-to-gage correlation. A

decrease of the IR threshold, on the other hand, would diminish not only bias and MAE, but also the satellite-to-gage correlation, especially over the mountains, where warm rain is more frequent. Overall, the sensitivity to the IR threshold is not very strong.

The sensitivity to the VIS threshold is even less pronounced. A change of five percentage points in the albedo threshold causes a change in relative bias of just two percentage points and no relevant change in other statistics.

7.3.6 Analytical Distributions

Analytical distributions are derived by fitting the $F_R(0)$, μ_{LNR}, and σ_{LNR} parameters to the sample distributions determined with the same procedure described in Section 7.3.1, and increasing the minimum number of samples used to determine the empirical cdfs to 150. This increase improves the reliability of the analytical distributions, but comes at the cost of a lower resolution in the IR/VIS/WV space. Further, the precipitation generated by sampling the analytical distributions is limited to being below 1.25 times the maximum observed precipitation. This is done to avoid the possibility that exceedingly high precipitation rates are generated because of misestimating the distribution statistics. Finally, precipitation rates below 0.5 mm h^{-1} are considered negligible.

Table 7.10 and Table 7.11 show that the analytical distribution model (AIVWNUU) has a lower bias and lower MAE than the empirical model and a slightly lower satellite-to-gage correlation. This is probably due to the combined effect of a coarser IR/VIS/WV-precipitation relation and to the approximations in the fitting of the sample distributions. All other statistics are very close to those produced by EIVWNUU, for both the calibration dataset and the validation dataset. Figure 7.3 shows that AIVWNUU's bias is uniformly lower than EIVWNUU's bias. However, the satellite-to-gage correlation of monthly precipitation of the two procedures is practically the same

everywhere. Further, both models have a bias of less than 40% at almost all gages, with the large majority below 20%.

Table 7.10 EIVWNUU and AIVWNUU performance for the calibration dataset (1996-1998).

Day

Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.04	0.13	0.45	0.57	1.10	14.9	0.91	0.02	0.36	0	0
AIVWNUU	-0.01	0.12	0.44	0.57	1.08	14.9	0.91	0.02	0.37	0	0

Dekad

Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.04	0.13	0.65	0.72	0.55	64.8	0.80	0.12	0.21	27	40
AIVWNUU	-0.01	0.12	0.64	0.72	0.54	63.8	0.79	0.12	0.21	28	43

Month

Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.04	0.13	0.74	0.75	0.39	121.7	0.74	0.16	0.24	22	42
AIVWNUU	-0.01	0.12	0.74	0.75	0.38	119.6	0.74	0.14	0.23	30	39

Table 7.11 EIVWNUU and AIVWNUU performance for the validation dataset (1996-1997).

Day

Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.08	0.19	0.47	0.55	1.14	13.3	0.91	0.03	0.43	0	0
AIVWNUU	0.02	0.17	0.47	0.54	1.12	13.1	0.91	0.03	0.44	0	0

Dekad

Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.08	0.19	0.70	0.76	0.60	60.1	0.77	0.15	0.25	17	28
AIVWNUU	0.02	0.17	0.70	0.76	0.58	58.6	0.77	0.14	0.26	17	24

Month

Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.08	0.19	0.79	0.81	0.43	112.3	0.69	0.21	0.27	19	32
AIVWNUU	0.02	0.17	0.79	0.81	0.42	109.7	0.69	0.19	0.26	19	32

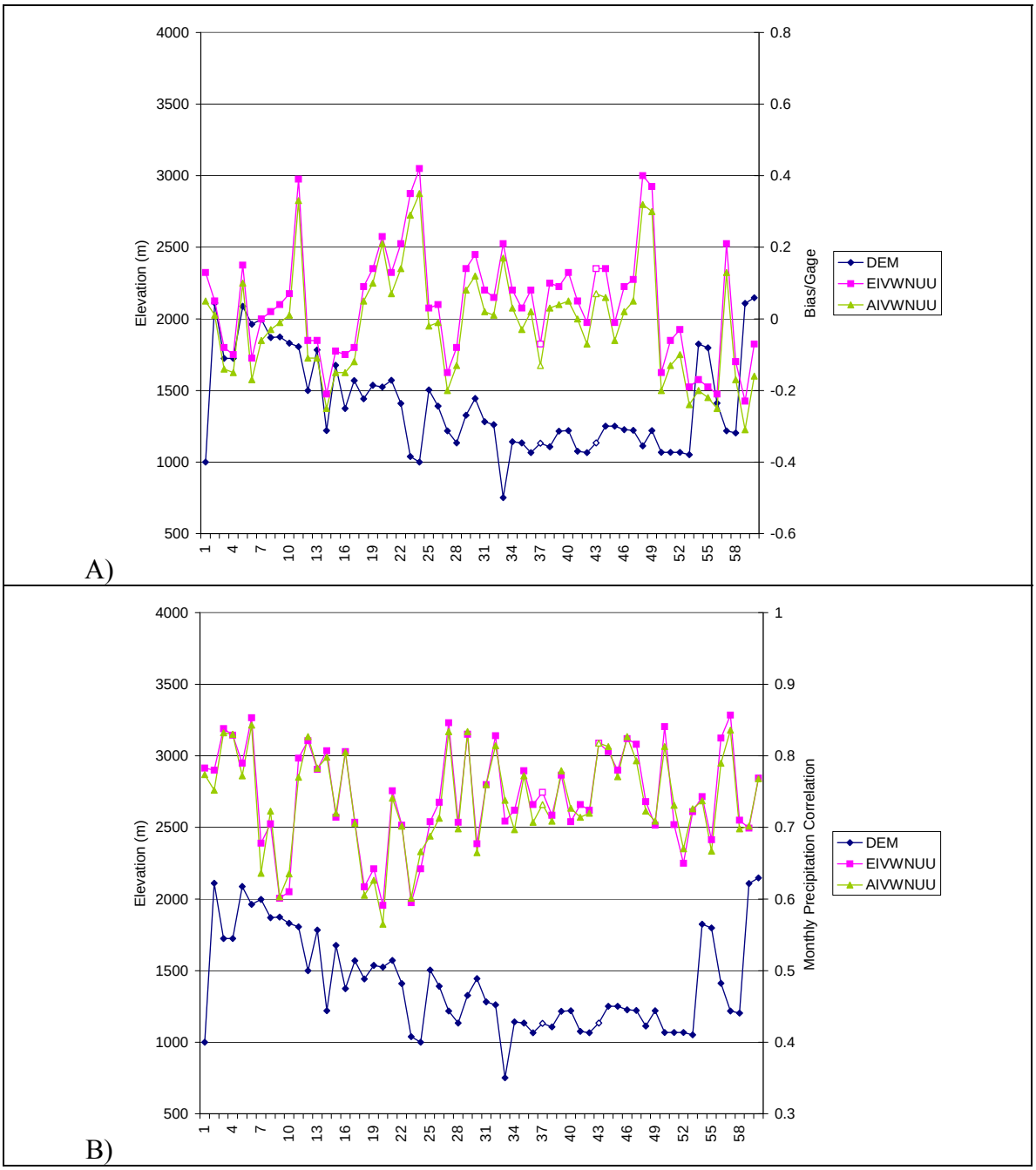


Figure 7.3 EIVWNUU and AIVWNUU behavior for the calibration dataset. A) Bias/Gage; B) Satellite-to-Gage correlation of monthly precipitation. Lake gage marks have white interior.

7.3.7 Discussion

The correlation between EIVWNUU and rain gages at the daily level (0.45) seems low, but the following factors should be considered:

1. The single station correlation varies between 0.29 and 0.61.
2. There is a relevant difference in the spatial scale of the gage (< 1 m^2) and satellite (>25 km^2) measured precipitation. The shorter the temporal resolution of the precipitation estimate, the higher the expected impact of this difference. This is especially true for the tropics, where precipitation is highly variable.
3. In several parts of the basin, geographical features (mountains, ridges, valleys, and shorelines) strongly influence precipitation patterns, even at the sub-pixel level;
4. Daily rain gage data were often of poor quality. It is very likely that quality control has eliminated the most problematic gages, but it is also almost certain that errors still affect several gages, limiting the model's performance.

Figure 7.4 may help one better appreciate the daily correlation between rain gage and satellite-based precipitation estimates. Gages 10046 and 1000419 are the only two instances in the dataset for which the closest neighboring gage is located within the same pixel. The correlations of these two stations with EIVWNUU are 0.53 and 0.50 respectively, while the correlations of the gages with their closest neighbor are 0.64 and 0.77 respectively.

Overall, EIVWNUU average estimation tracks gage data better than GPI and 3B42, both in terms of bias and in terms of correlation. The performance of the average estimation shows good consistency over the entire calibration set (1996-1998), a subset of it (1998), and the validation set (1996-1997), indicating that the method is sufficiently robust. This is especially notable considering that (1) no TRMM data is available for 1996-1997, (2) the visible data were derived according to two different methodologies in

1996-1997 and 1998-1999 (Eumetsat, 2004; Rigolier et al., 2002), and (3) Meteosat changed satellite platforms in 1996 and 1998. Further, there is good potential to improve the estimator performance using more TRMM data.

Similar considerations can be made about AIVWNUU, the behavior of which closely resembles that of EIVWNUU.

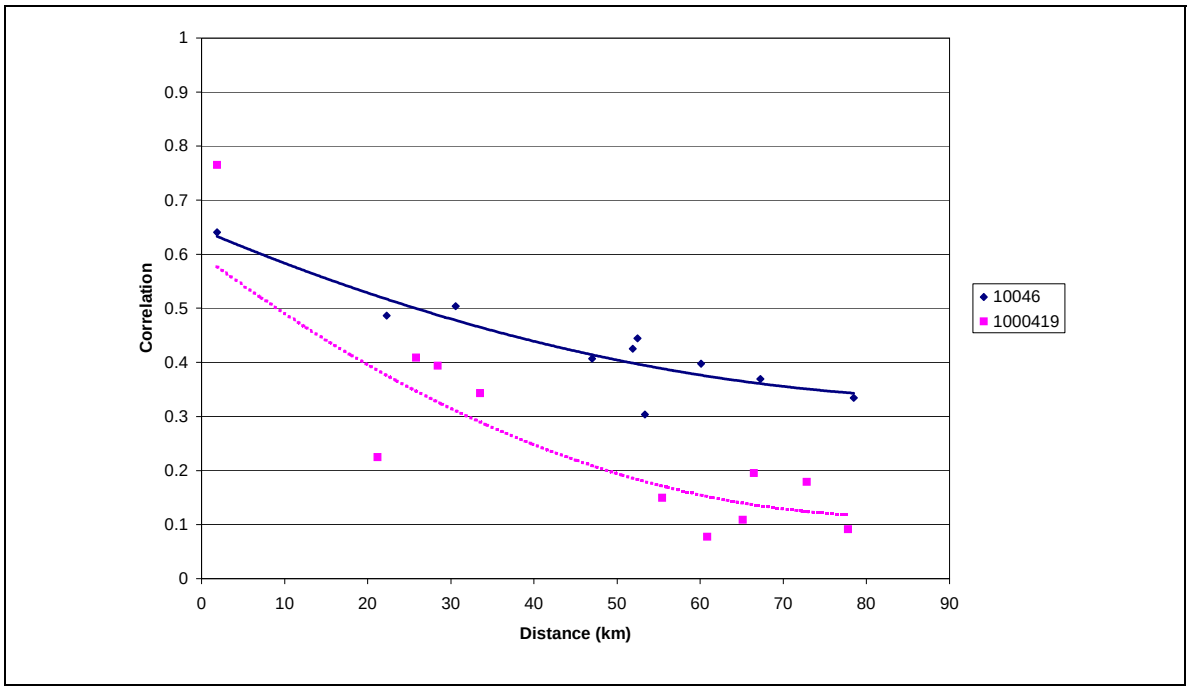


Figure 7.4 Rain gage daily precipitation correlation of as a function of gage distance (correlation EIVWNUU-10046 =0.53, EIVWNUU -1000419=0.50).

By contrast, the representation of precipitation variability is not satisfying. The 95% compliance rate can be considered acceptable at the daily level (91%), but not at the dekad (79%) and monthly resolutions (72%).

Extensive investigation has shown that the reason the model is able to track the precipitation average well, but not account for the full extent of precipitation variability can be attributed to the fact that precipitation rates at consecutive timeslots are chosen independently, not taking into account the temporal autocorrelation that may be present.

TRMM images do not directly provide a measure of the precipitation temporal autocorrelation. In order to test this hyporesearch, the precipitation temporal autocorrelation was derived from the spatial TRMM images considering the precipitation correlation in adjacent pixels at stage *t* and *t-1* as determined by the neural network (Figure 7.5).

The precipitation temporal autocorrelation is very significant especially at the later stages, where stratiform or mixed precipitation is dominant. It is also much stronger than the autocorrelation in the precipitation produced by the EIVWNUU model (the "Ensemble" series of Figure 7.5).

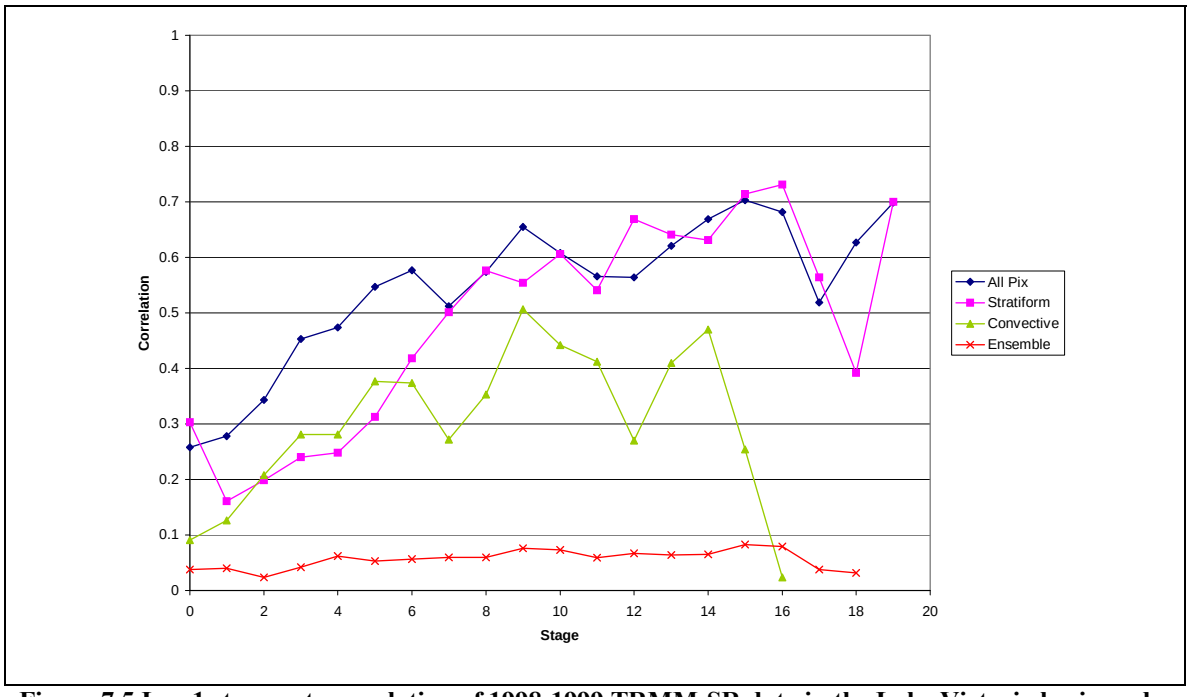


Figure 7.5 Lag 1 stage autocorrelation of 1998-1999 TRMM-SR data in the Lake Victoria basin and in the EIVWNUU model.

7.4 Estimation Incorporating Precipitation Autocorrelation

The most typical way to model the precipitation temporal correlation is through the Markov chain model. In this model, the future state of the process (time *t+1*) is influenced only by the present state (time *t*) and not by the anterior states, that is:

$$P\{X_{t+1}= x_{t+1} \mid X_t= x_t, X_{t-1} = x_{t-1}, \ldots, X_0= x_0\} = P\{ X_{t+1}= x_{t+1} \mid X_t= x_t \} \qquad (7.1)$$

This model is possibly a first approximation of the real temporal dependency of precipitation, but it is adequate for the stage of this research and compatible with the data available for the Lake Victoria basin.

7.4.1 Empirical Distributions

The one-step autocorrelation of single-pixel precipitation is implemented by conditioning the rain-rate distribution on precipitation at time *t-1* in addition to IR, VIS/WV, stage, orography, and month. That is, for each member *X* of the ensemble:

$$\mathbf{precip(t)} = \begin{cases} f[\mathrm{IR(t)}, \mathrm{VIS(t)/WV(t)}, \mathrm{stage(t)}, \mathbf{precip}(t-1), \mathrm{orography}, \mathrm{month}, \varepsilon] & \mathrm{stage(t)} \geq 0 \\ f[\mathrm{IR(t)}, \mathrm{VIS(t)/WV(t)}, \mathrm{stage(t)}, \mathrm{orography}, \mathrm{month}, \varepsilon] & \mathrm{stage(t)} < 0 \end{cases} \qquad (7.2)$$

where: precip(t) = precipitation rate at time *t* for element *X*;

precip(t-1) = precipitation rate at time *t-1* for element *X*;

IR(t) = IR at time *t*;

VIS(t) = VIS at time *t*;

WV(t) = WV at time *t*;

stage(t) = storm stage at time *t*;

ε =random error;

The derivation of the temporal precipitation relation from TRMM data is possible only for the timeslots belonging to a convective storm. The precipitation over pixels not belonging to a convective storm event (identified by the ANN) is presumed to be temporally uncorrelated.

The rational of this approach is that the new input variable (*precip(t-1)*) restricts the pool of available precipitation rates that can be used to build the rain-rate distributions for time t, thus establishing a relation between precipitation at time t and at time *t-1*. At the same time, an entire set of possible precipitation rates is associated with *precip(t-1)*, thus directly accounting for the random component of this relation. Coherently with the empirical distribution approach, this implementation of the temporal correlation is not based on any model of the probabilistic distribution of the instantaneous rain-rates, something that is very suitable to the markedly skewed distribution of TRMM SR data.

Adding *precip(t-1)* to the input variable set further fragments the data available for building the rain-rate distributions and prevents the use of *VIS(t)/WV(t)* and month information during convective storms . Consequently, the precipitation at time t is determined according to the following formula:

$$\mathbf{precip(t)} = \begin{cases} f[\mathrm{IR(t)}, \mathrm{stage(t)}, \mathbf{precip}(\mathrm{t-1}), \mathrm{orography}, \varepsilon] & \mathrm{stage(t)} \geq 0 \\ f[\mathrm{IR(t)}, \mathrm{VIS(t)/WV(t)}, \mathrm{stage(t)}, \mathrm{orography}, \mathrm{month}, \varepsilon] & \mathrm{stage(t)} < 0 \end{cases} \tag{7.3}$$

where: precip(t) = precipitation rate at time *t* for element *X*;

precip(t-1) = precipitation rate at time *t-1* for element *X*;

IR(t) = IR at time t;

VIS(t) = VIS at time t;

WV(t) =WV at time t;

stage(t) = storm stage at time *t*;

ε = random error;

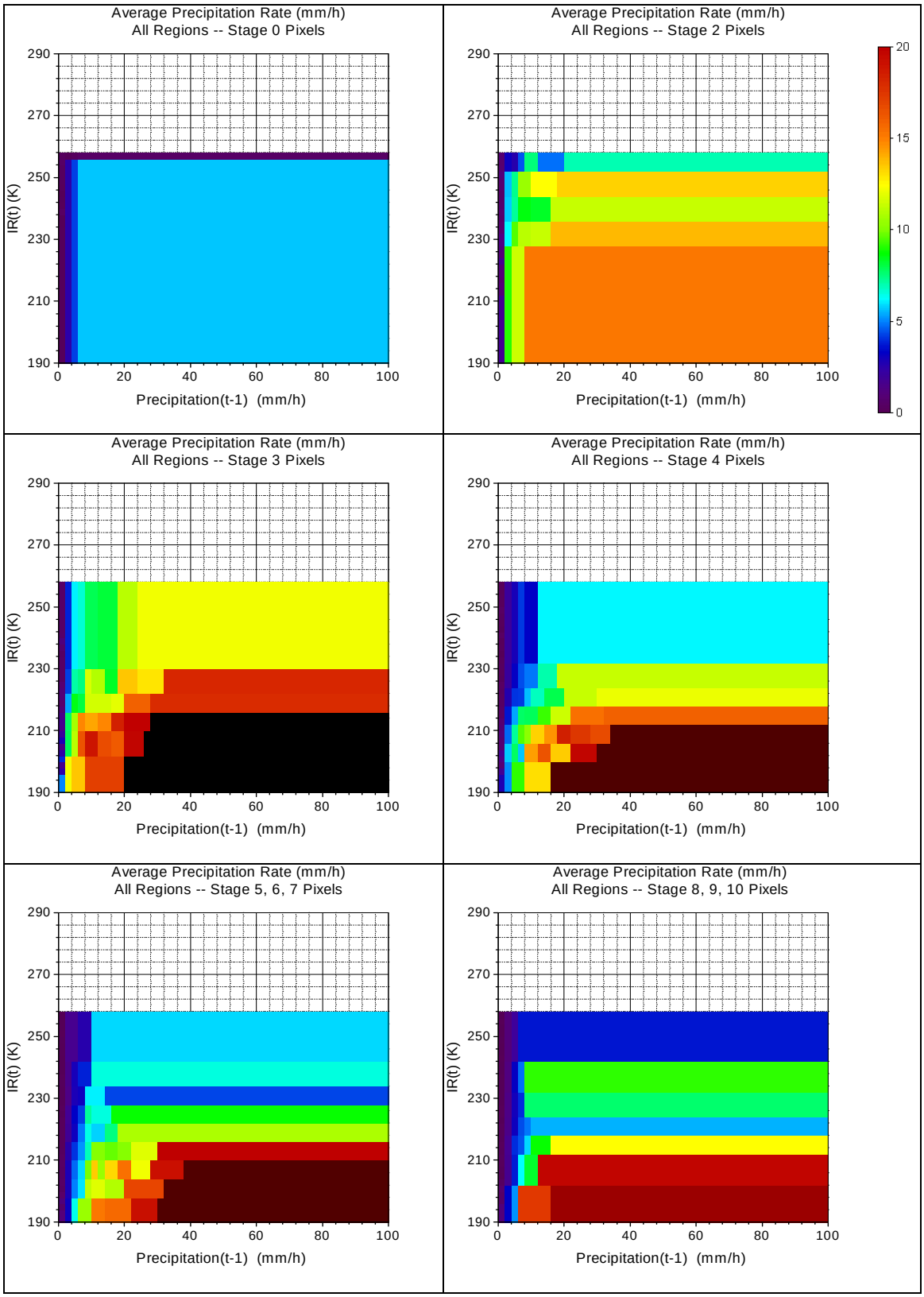


Figure 7.6 Average precipitation rate as a function of IR and precip(t-1) for all terrains and at least 75 TRMM data per bin.

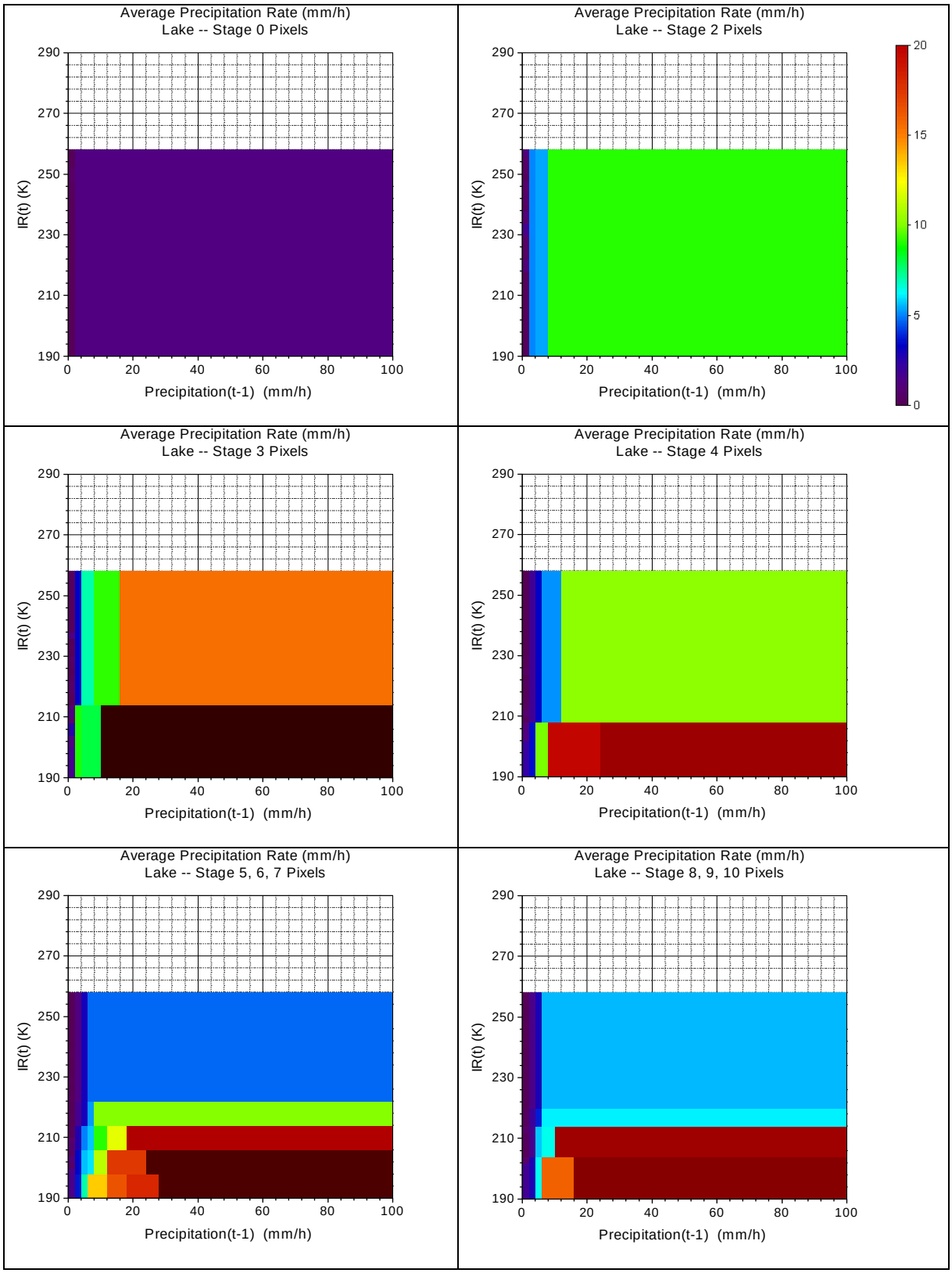


Figure 7.7 Average precipitation rate as a function of IR and precip(t-1) for lake pixels and at least 75 TRMM data per bin.

The average of the precipitation distributions used to determine *precip(t)* (Figure 7.6 and Figure 7.7) behaves as expected: the average rate increases with decreasing IR and with increasing *precip(t-1)*. Further, the average rain rate peaks between stages 3 and 7. The segmentation of the *IR(t)/precip(t-1)* space looks very rough for *stage(t)*=0, but this is actually due to the fact that most of the *precip(t-1)* data in this case is zero. The following stages are characterized by a larger variety of *precip(t-1*), so that the resulting rain-rate map is more finely detailed.

Figure 7.7 shows the effect that the subdivision of TRMM PR data according to the orography may have on the average rain-rate mapping and, consequently, on the rain-rate distributions: the lake pixels are barely enough to guarantee an acceptably detailed mapping of the precipitation. For example, precipitation maps for stage 0 to stage 1 and *precip(t-1)>0* are insensitive to IR. The future availability of more TRMM PR under this project would enable a more accurate estimation of these relationships.

7.4.1.1 Autocorrelated Model Calibration (1996-1998)

Table 7.12 reveals that the autocorrelated estimator (EIVWNCU) has a bias similar in magnitude to that of the uncorrelated estimator (EIVWNUU), but a slightly lower absolute bias. On the other hand, its correlation with rain gage data is lower than that of the EIVWNUU, especially at the monthly level. This is the effect of not including VIS/WV and month in the variable used to identify the precipitation distribution to be used during storm events.

On the other hand, the 95%C and KS indexes improve substantially at all levels. The 95%C reaches 94%, 91.4%, and 88.3% at the daily, dekad, and monthly resolutions. The number of stations passing the KS test at the dekad and monthly level increases, especially at the 10% significance level. Despite all these improvements in the representation of the precipitation variability, the KS test at the daily level still indicates that no station has a gage percentile distribution close to the uniform distribution.

Table 7.12 EIVWNUU and EIVWNCU performance for the calibration dataset (1996-1998).

Day

Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.04	0.13	0.45	0.57	1.10	14.9	0.91	0.02	0.36	0	0
EIVWNCU	-0.03	0.12	0.43	0.55	1.08	20.6	0.94	0.00	0.38	0	0

Dekad

Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.04	0.13	0.65	0.72	0.55	64.8	0.80	0.12	0.21	27	40
EIVWNCU	-0.03	0.12	0.63	0.71	0.55	91.5	0.91	0.04	0.18	37	44

Month

Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.04	0.13	0.74	0.75	0.39	121.7	0.74	0.16	0.24	22	42
EIVWNCU	-0.03	0.12	0.71	0.74	0.39	175.2	0.88	0.06	0.18	41	52

The most likely explanation of this behavior is that the distribution of daily precipitation is not continuous, but presents a discrete probability for precipitation equal to zero, posing a problem on how to compute the associated percentile. The definition used here, as suggested in several sources, is:

$$\text{Percentile(gage)} = \text{Prob(satellite<gage)} \tag{7.4}$$

where gage = precipitation recorded at the gage;

satellite = precipitation estimate (ensemble).

If the weather is warm and clear during an entire day, it is most likely that all elements of the daily precipitation ensemble are zero. If it happens that the gage records no rain, the associated percentile is 0.998 for a 500-member ensemble. In the continuous distribution case, for which the KS test is formally applicable, this percentile is associated with the event that gage precipitation is higher than all ensemble elements, a completely different situation. The number of such events in this region is much higher than 0.5% as

in certain parts of the basin no rain is recorded for three to four months. In the dekad and monthly distributions, on the other hand, such a high discrete probability is not present, making these cases more suitable for the KS test.

Given that the uncorrelated dekad and monthly performances improved when "monthly" precipitation distributions were used, there is a good chance that these performances will further improve when additional TRMM PR data are used.

Figure 7.8 shows that adding *precip(t-1)* to the criteria used for selecting *precip(t)* successfully increases the autocorrelation of the ensemble sequences to a level closer to that determined from the TRMM images.

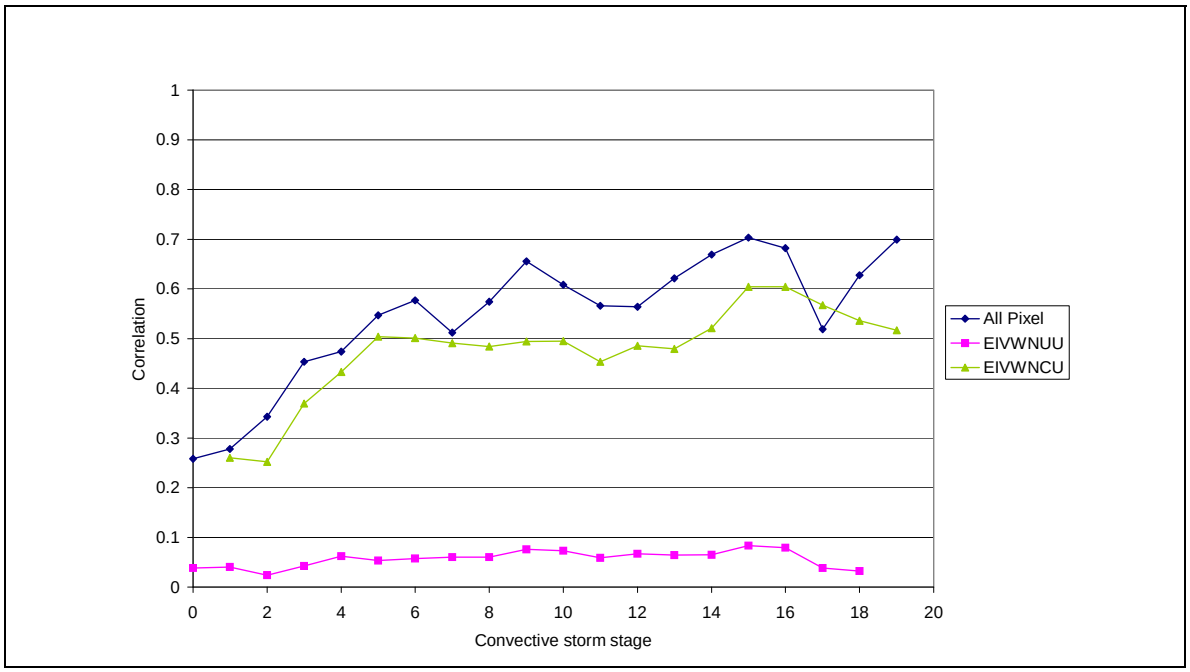


Figure 7.8 One-step autocorrelation of convective storm precipitation determined directly from the TRMM data (All Pixel) and as reproduced by the EIVWNCU and EIVWNUU models.

7.4.1.2 Autocorrelated Model Validation (1996-1997)

The EIVWNCU model performance for the validation data set is similar to that for the calibration set. As with EIVWNUU, the bias is slightly higher, most likely because the southern part of the basin, which was practically absent from the calibration set, is less rainy. EIVWNCU correlation is also higher than in the calibration dataset, similarly to EIVWNUU. The 95% compliance rate is a couple of points lower than it is in the calibration set, probably because of the bias increase.

Overall, the absolute model performance is very similar to that of the calibration set and exhibits the relative differences with EIVWNUU.

Table 7.13 EIVWNUU and EIVWNCU performance for the validation dataset (1996-1997).

Day Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.08	0.19	0.47	0.55	1.14	13.3	0.91	0.03	0.43	0	0
EIVWNCU	0.03	0.16	0.45	0.54	1.14	18.7	0.95	0.00	0.43	0	0
Dekad Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.08	0.19	0.70	0.76	0.60	60.1	0.77	0.15	0.25	17	28
EIVWNCU	0.03	0.16	0.66	0.75	0.61	87.3	0.89	0.06	0.23	20	28
Month Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNUU	0.08	0.19	0.79	0.81	0.43	112.3	0.69	0.21	0.27	19	32
EIVWNCU	0.03	0.16	0.76	0.79	0.44	165.6	0.86	0.08	0.18	35	40

7.4.2 Analytical Distributions

As mentioned in Section 6.4, one of the advantages of the analytical representation of cdf is that it allows to represent the spatial and temporal correlation by operating on the analytical model parameters instead of directly on the cdf. In the case of

the temporal correlation, this operation is normally done by establishing a linear relation between the distribution's parameters at time *t* and at time *t+1*. For the mixed continuous-discrete model of the precipitation distribution in this research (equation 6.1), this implies the need to establish two distinct relations, one for the discrete probability of rain/no-rain, and one for the distribution of positive rain intensity.

For any point *x* and a time *t*, the no-rain event can be modeled as a binomial random variable, called indicator *i(x,t,0)*, with outcome 1 if there is no rain over *x* at time *t* and 0 if there is. The unconditional probability of no-rain at any time *t*, *p(x,t,0)*, can be estimated from satellite information (IR(t), VIS(t)/WV(t), and stage(t)), the month, and the terrain. Furthermore, the following property holds for indicator variables:

$$E\{i(x, t,0)\}=1*p(x,t,0)+0*(1-p(x,t,0)) = p(x,t,0) \quad (7.5)$$

The probability of no-rain conditioned by the outcome of the rain/no-rain event at *t-1*, can then be estimated using the following regression model:

$$[i(x,t,0)]^* = p(x,t,0)+a(x,t)(i(x, t-1,0)-p(x,t-1,0)) \quad (7.6)$$

where: *i(x,t,0)* is the no-rain indicator variable for *x* at time *t*;

i(x,t-1,0) is the no-rain indicator variable for *x* at time *t-1*;

p(x,t,0) is the unconditional probability of rain for *x* at time *t*;

p(x,t-1,0) is the unconditional probability of rain for *x* at time *t-1*;

a(x,t) is a regression coefficient.

Equation 7.6 states that the conditional probability of no-rain at pixel *x* and time *t* is given by the unconditional probability of no-rain plus the weighted difference between the actual probability that no-rain occurred at time *t-1* and the respective unconditional

probability. The expectation of the estimator *[i(x,t,0)]** is *p(x, t, 0)* by virtue of equation 7.6. The value of the coefficient *a(x,t)* can be determined by minimizing the variance of the estimation error, that is

$$\min_{a(x,t)} E\left\{(i(x,t,0)^* - i(x,t,0))^2\right\} \tag{7.7}$$

This is obtained by setting the first derivative of 7.7 with respect to *a(x,t)* to zero:

$$E\left\{(p(x,t,0) + a(x,t)(i(x,t-1,0) - p(x,t-1,0)) - i(x,t,0))(i(x,t-1,0) - p(x,t-1,0))\right\} = 0 \tag{7.8}$$

which results in

$$a(x,t) = \frac{E\{i(x,t,0)i(x,t-1,0)\} - p(x,t,0)\,p(x,t-1,0)}{E\{i(x,t-1,0)i(x,t-1,0)\} - p(x,t-1,0)\,p(x,t-1,0)} = \frac{Cov_I(x,t,t-1)}{Cov_I(x,t-1,t-1)} \tag{7.9}$$

where: $Cov_I\{x,t,t-1\}$ is the covariance of *i(x,t,0)* and *i(x,t-1,0)*;

$Cov_I\{x,t-1,t-1\}$ is the covariance of *i(x,t-1,0)* and *i(x,t-1,0)*, which is also equal to the variance of *i(x,t-1,0)*.

$$[i(x,t,0)]^* = p(x,t,0) + \frac{Cov_I(x,t,t-1)}{Cov_I(x,t-1,t-1)}(i(x,t-1,0) - p(x,t-1,0)) \tag{7.10}$$

The value of $Cov_I(x,t,t-1)$ is supposed to be independent of *x* and *t*, but dependent on stage(t), making *a(x,t)* also a function of stage(t) only. This value can be derived from the TRMM PR images with the same technique adopted in Sections 7.3.7 and 7.4.1 for $Cov_R(t,t-1)$ and the precipitation autocorrelated empirical distributions. Figure 7.9 shows that the value of *a(stage)* is high for the pixels belonging to a convective storm and low

for pixels not belonging to a convective storm. Within the storm, the regression coefficient seems to rise with the duration of the storm until stage 13. Beyond this stage, the wildly fluctuating values of *a(stage)* are likely due to the small number of data points used to compile the statistics. Thus, it was decided to use only three coefficients (Table 7.14).

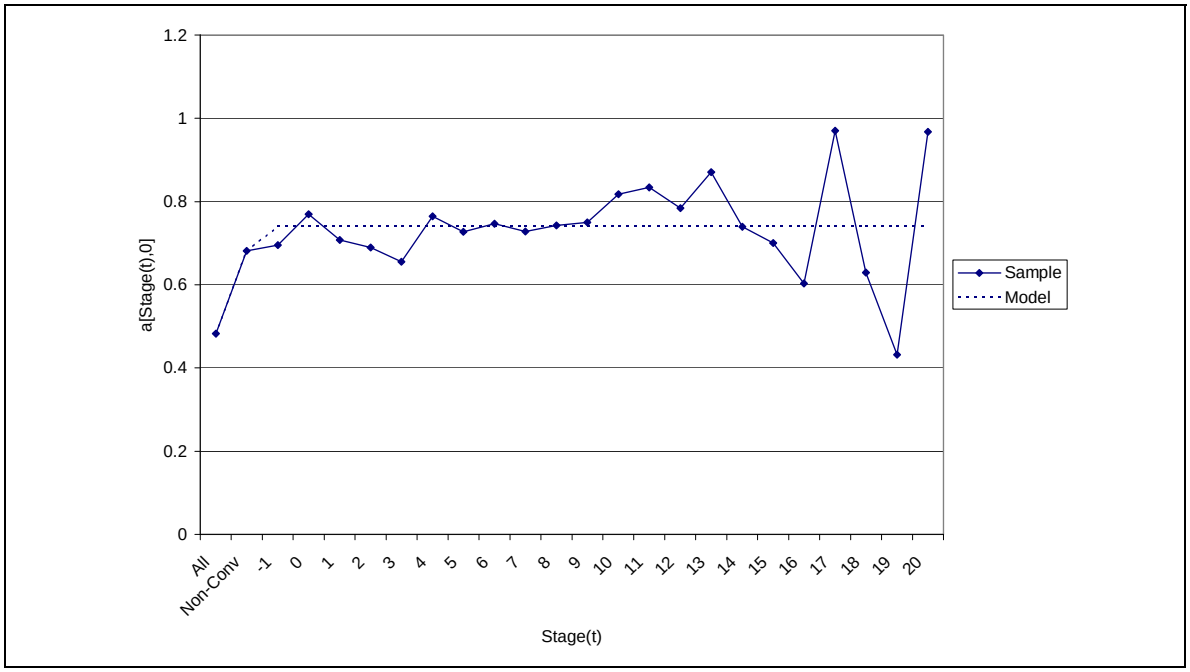


Figure 7.9 Regression coefficient for implementing the 1-lag temporal correlation of the probability of no rain.

Table 7.14 Probability of rain 1-lag regression coefficient.

Stage	Non Classifiable	Non-convective	-1 to 20
a(Stage)	0.483	0.681	0.740

A similar approach is taken also to implement the temporal correlation of the conditional rain rates. In this case, however, the random variable is continuous and

lognormally distributed. Therefore, for any point *x* and a time *t*, the logarithm of the conditional precipitation rate can be modeled as a normal random variable, say *Z(x,t)*. The unconditional mean and variance of this variable, $\mu_{LNR}(x,t)$ and $\sigma^2_{LNR}(x,t)$, can be estimated from satellite information (IR(t), VIS(t)/WV(t), and stage(t)), the month, and the terrain.

The mean of *Z(x,t)* conditioned by the value of the logarithm of the precipitation rate at *t-1*, *z(x,t-1)*, can then be estimated using the following regression model:

$$E\{Z(x,t) \mid z(x,t)\} = \mu_{LNR}(x,t) + b(x,t)[z(x,t-1) - \mu_{LNR}(x,t-1)] \tag{7.11}$$

The expectation of this estimator is by definition equal to $\mu_{LNR}(x,t)$. Using a procedure similar to equations 7.8 to 7.10, the estimators of the conditional mean and variance given *z(x, t-1)* is obtained as:

$$\begin{cases} \mu_{LNR}(x,t \mid z(x,t-1))^* = m_{LNR}(x,t) + \dfrac{C_{LNR}(x,t,t-1)}{C_{LNR}(x,t-1,t-1)}[z(x,t-1) - \mu_{LNR}(x,t-1)] \\ \sigma^2{}_{LNR}(x,t,t \mid z(x,t-1)) = \sigma_{LNR}(x,t)(1 - \rho_{LNR}{}^2(x,t,t-1)) \end{cases} \tag{7.12}$$

The value of $Cov_{LNR}(x,t,t-1)$ is independent of *x* and *t*, but dependent on stage(t), making *a(x,t)* and $\rho_{LNR}(x,t,t-1)$also functions of stage(t) only. The two coefficients can be derived from the TRMM PR images with the same technique adopted previously for $Cov_I(x,t,t-1)$. The number of neighboring rainy pixels at consecutive stages is small, introducing irregularities in $\rho_{LNR}(stage)$ (Figure 7.10). In view of this, only three coefficients are selected (Table 7.15). The distribution of the *a(stage)* is even more erratic. Therefore, equation 7.12 was slightly changed, setting *a(stage)* equal to the more stable $\rho_{LNR}(stage)$.

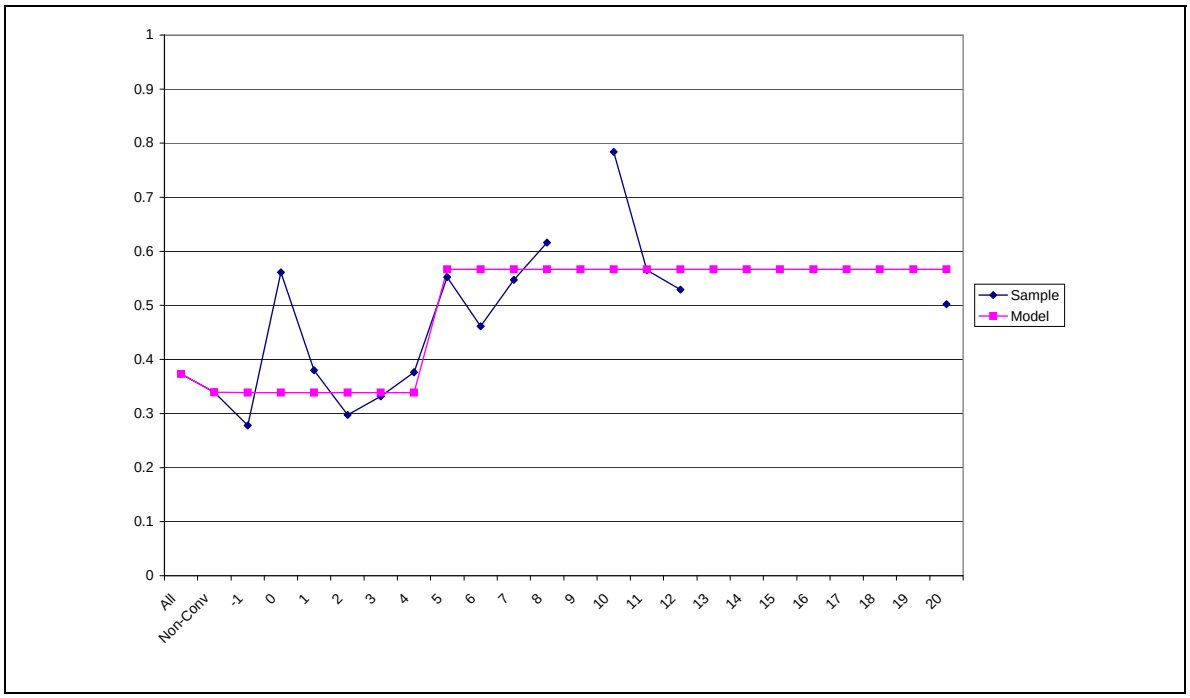


Figure 7.10 Correlation coefficient ρ_{LNR}(stage(t))

Table 7.15 Correlation coefficient of the logarithm of the conditional precipitation.

Stage	Non Classifiable	Non-convective, -1 to 5	-1 to 20
ρ_{LNR}*(stage)*	0.373	0.340	0.567

7.4.2.1 Autocorrelated Model Calibration (1996-1998)

The temporally uncorrelated analytical model AIVWNUU described in Section 7.3.6 was changed by adding the temporal autocorrelation procedure described earlier with the parameters listed in Table 7.14 and Table 7.15. The results in estimating the calibration set showed that the model performed well, but had a slightly higher bias (~6%). Further, the 95% compliance rate at the monthly level was one percentage point smaller than that yielded by the empirical model. To compensate for these two effects,

the screening of low precipitation was increased to 1.0 mm h^{-1} and the coefficients in Table 7.15 were gradually increased to 15% because of their low and uncertain value. The coefficients in Table 7.14 were considered comparatively stronger and more reliable, so they were not changed.

With these changes, performances of the temporally autocorrelated estimator with analytical distributions, AIVWNCU, looked very close to those of EIVWNCU (Table 7.16). One of the main differences between the two models is the higher satellite-to-gage correlation of AIVWNCU, especially at the monthly resolution (Figure 7.11). This is because AIVWNCU's parameters are always functions of all satellite and month variables, while only a limited set of the input variables are used in EIVWNCU during the convective storms.

Further, AIVWNCU features smaller 95% confidence intervals. Although this does not seem to affect the compliance rate in the calibration set, it may result in reduced reliability.

Table 7.16 EIVWNCU and AIVWNCU performance for the calibration dataset (1996-1998).

Day Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNCU	-0.03	0.12	0.43	0.55	1.08	20.6	0.94	0.00	0.38	0	0
AIVWNCU	0.01	0.12	0.44	0.56	1.09	19.7	0.94	0.00	0.39	0	0
Dekad Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNCU	-0.03	0.12	0.63	0.71	0.55	91.5	0.91	0.04	0.18	37	44
AIVWNCU	0.01	0.12	0.64	0.72	0.55	88.2	0.92	0.03	0.17	37	45
Month Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNCU	-0.03	0.12	0.71	0.74	0.39	175.2	0.88	0.06	0.18	41	52
AIVWNCU	0.01	0.12	0.73	0.75	0.39	167.8	0.88	0.06	0.18	40	52

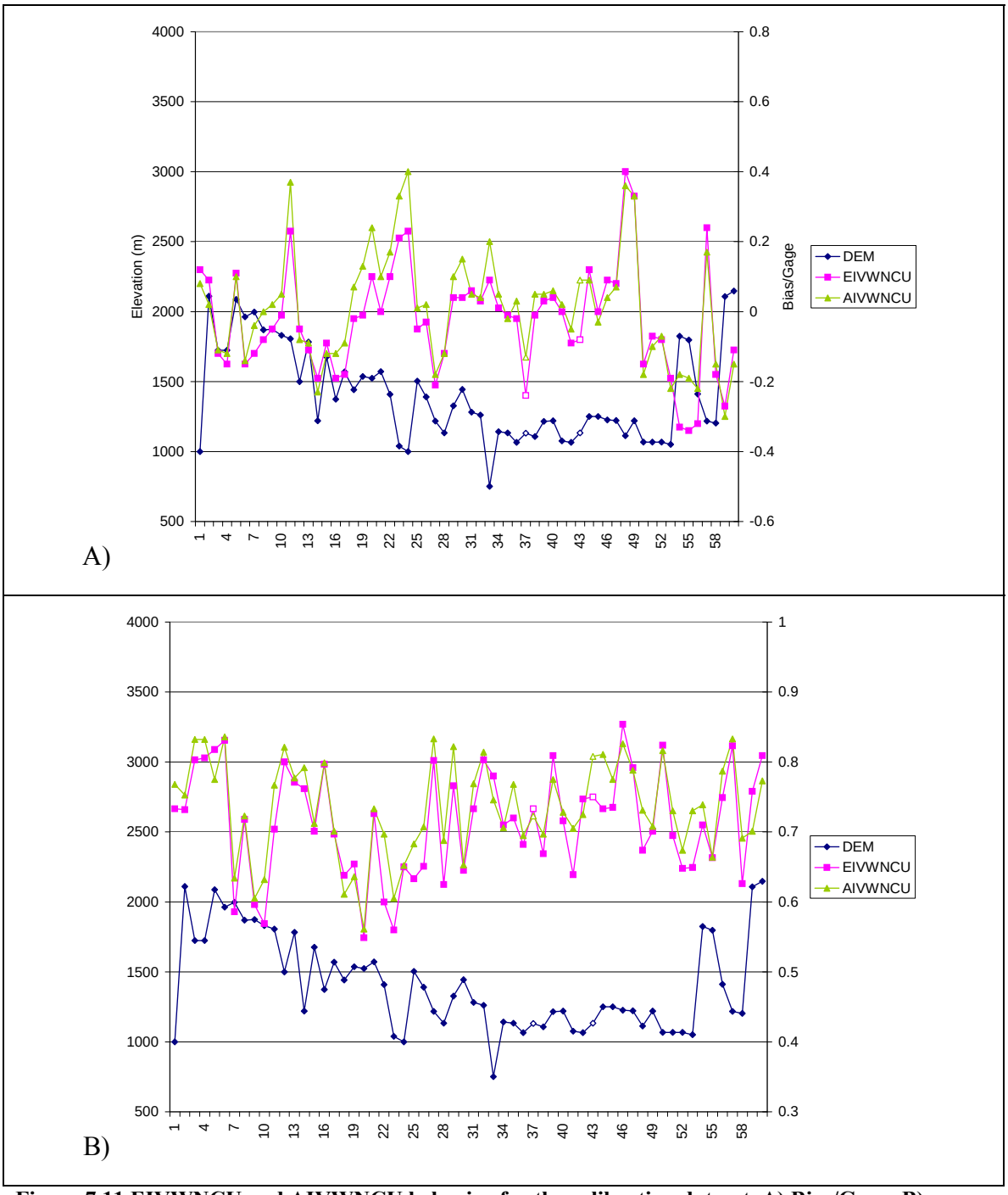


Figure 7.11 EIVWNCU and AIVWNCU behavior for the calibration dataset. A) Bias/Gage; B) Satellite-to-Gage correlation of monthly precipitation. Lake gage marks have white interior.

A comparisons of Figure 7.3 and Figure 7.11 shows that AIVWNCU bias increases significantly for pixels below 1000 m. This is probably due to the low number of samples at this elevation, which makes the IR/VIS and IR/WV relations very rough. This problem is exaggerated even more by the higher number of elements used by the analytical models to compute their distributions.

7.4.2.2 Autocorrelated Model Validation (1996-1997)

AIVWNCU's estimates have a slightly higher bias during the validation period, but a remarkably higher absolute bias and MAE, a sign that large underestimations are compensated for by large overestimations. This problem also affects EIVWNCU, but to a smaller extent.

Table 7.17 EIVWNCU and AIVWNCU performance for the validation dataset (1996-1997).

Day Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNCU	0.03	0.16	0.45	0.54	1.14	18.7	0.95	0.00	0.43	0	0
AIVWNCU	0.04	0.18	0.47	0.54	1.13	17.5	0.95	0.00	0.45	0	0
Dekad Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNCU	0.03	0.16	0.66	0.75	0.61	87.3	0.89	0.06	0.23	20	28
AIVWNCU	0.04	0.18	0.70	0.76	0.59	81.8	0.89	0.05	0.24	19	28
Month Method	Bias/ Gage	Abias /Gage	Corr	RCorr	MAE/ Gage	95%R	95%C	95%L	KS	KS_{10}	KS_{01}
EIVWNCU	0.03	0.16	0.76	0.79	0.44	165.6	0.86	0.08	0.18	35	40
AIVWNCU	0.04	0.18	0.78	0.81	0.43	154.5	0.84	0.09	0.21	28	38

AIVWNCU is demonstrated to have a higher gage-to-satellite correlation than EIVWNCU, but also a smaller 95% confidence interval, which, in this case, translates

into a lower reliability. One of the reasons for the worsening in performance is the presence in the validation dataset of two gages at low elevation.

7.5 Discussion

The previous sections showed that AIVWNCU and EIVWNCU have a very small average bias with respect to both the calibration and verification data sets. To fully understand the model capability, however, it is necessary to observe the distribution of the models' performance across the gage stations. Thus, the model performances for the entire 1996-1997 dataset are now considered, using all 97 gages to better represent the precipitation distribution over the region. Figure 7.12 A shows that EIVWNCU's absolute value of the relative bias is above 50% for only 1% of all stations and below 20% for 77% of the gages. For AIVWNCU, these figures are 2% and 76% respectively. Figure 7.12 B shows that EIVWNCU's satellite-to-gage monthly correlation is below 0.5 in only 4% of cases and above 0.7 in 69% of cases. For AIVWNCU, these two figures change to 1% and 76%. Figure 7.12C shows that EIVWNCU's monthly 95% compliance rate is below 0.8 in only 16% of the stations, while AIVWNCU's monthly 95% compliance is below 0.8 in 20%. Further, assuming that the two gages with the lowest compliance rate are outliers, EIVWNCU's average compliance rate increases to 89% (88% for AIVWNCU).

Figure 7.13sorts the gages from West to East and South to North, so it is very likely that adjacent gages are relatively close. Figure 7.13 A shows that there is no clear relation between bias and elevation. By contrast, monthly correlations reveal two areas in the western mountains where correlations are distinctly lower. However, since there are gages at high elevation where correlation is good, this does not seem to be a direct consequence of elevation, but most likely a consequence of a complex orography.

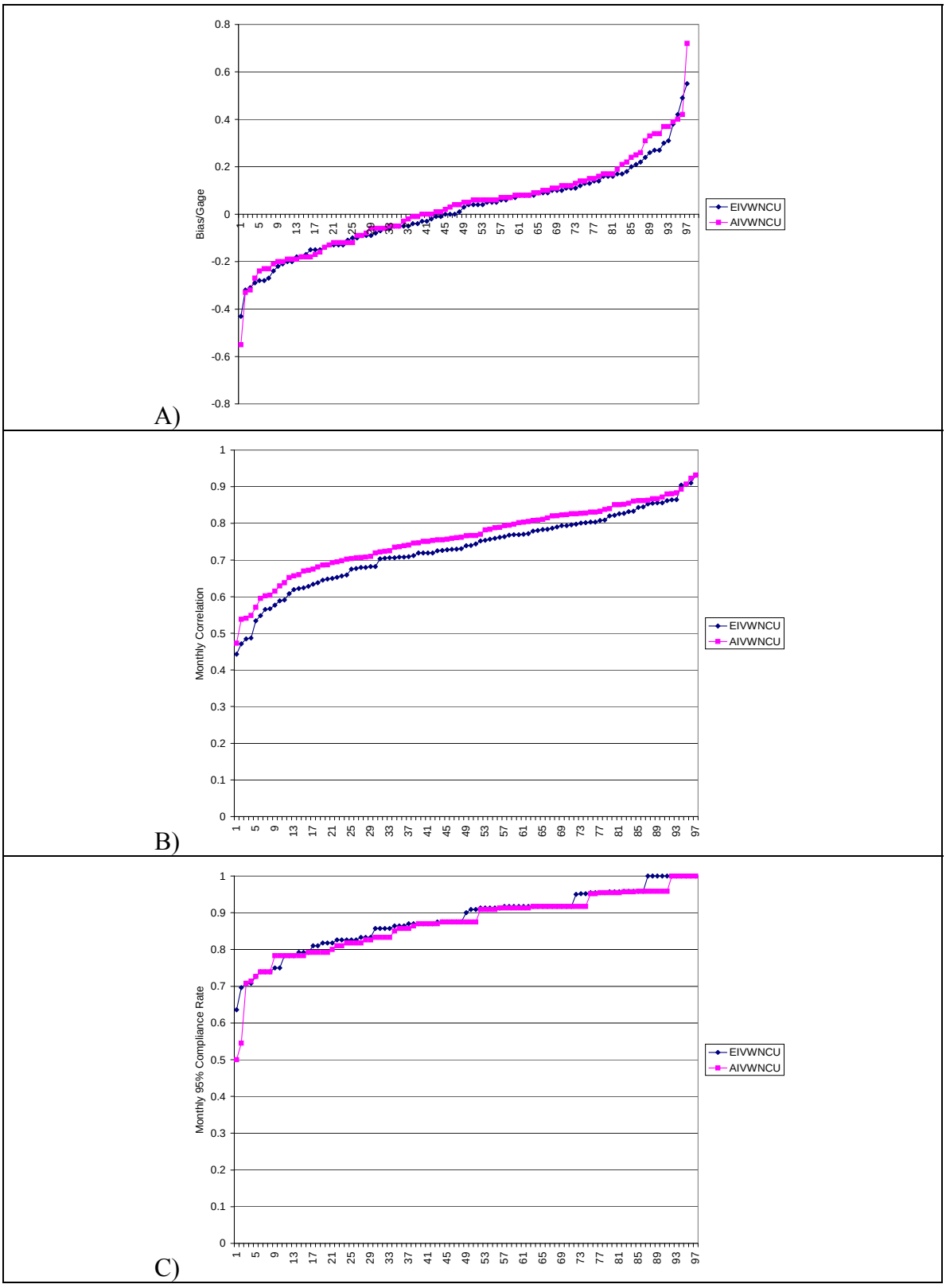


Figure 7.12 AIVWNCU and EIVWNCU relative bias, monthly correlation, and monthly 95% compliance rate distribution for the entire 1996-1997 dataset

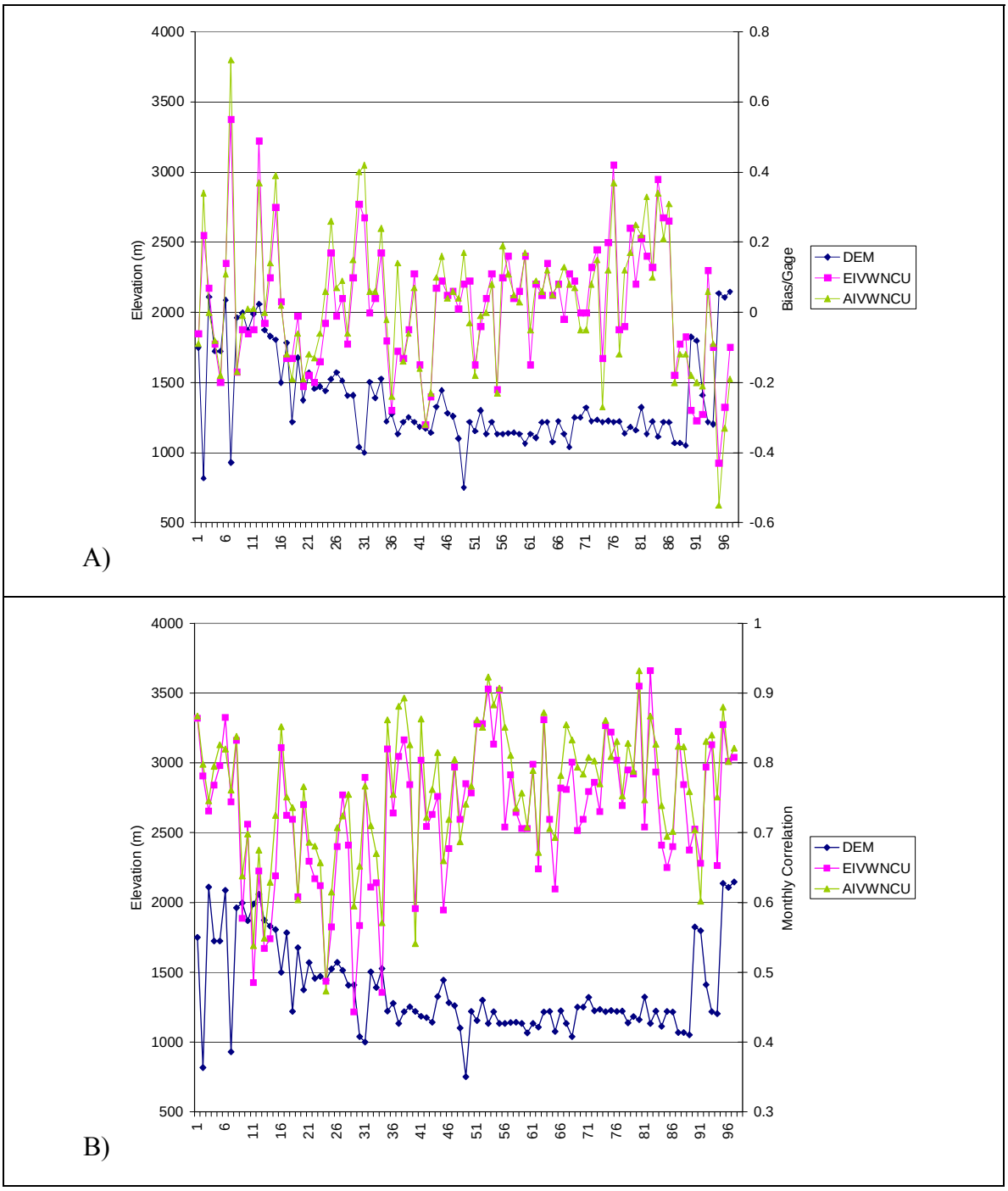


Figure 7.13 Relative bias (A) and monthly correlation (B) of the AIVWNCU and EIVWNCU models as a function of the gage elevation for the entire 1996-1997 dataset.

Figure 7.14 A shows that four of the gages for which EIVWNCU overestimates precipitation are situated in the Rift Valley, a deep and dry valley surrounded by very high mountains in the western side of the region.

Two other gages where EIVWNCU overestimates precipitation are located in the very orographically complex region of Mbarara in the western side of the basin. EIVWNCU seems to overestimate precipitation also in some areas shortly inland of the Lake Victoria shoreline. A possible explanation of this phenomenon, is that clouds carried by the lake breeze lose most of their humidity when they come over the shore because of convergence and are relatively dry when they move inland. EIVWNCU somewhat underestimates precipitation on the eastern mountains, probably due to a higher contribution of shallow precipitation of orographic origin in this area.

Figure 7.14 B shows that the gages with poor correlation are concentrated around some orographically complex regions in the western side of the basin and along the western shore of Lake Victoria. These observations, however, have to be re-evaluated after eliminating the possibility of questionable rain gage data quality.

The overall comparison between EIVWNCU and AIVWNCU shows that EIVWNCU is probably more robust than AIVWNCU. However, the difference in performances is not very large. Further, the assessment of AIVWNCU capabilities is hampered by the poor matching of gage precipitation at low elevations, due to the combination of the low density of data points in this elevation band and the higher data requirements set for AIVWNCU. It is very likely that these problems will be solved once more data are available. However, AIVWNCU's satellite-to-gage correlation and MAE are better than EIVWNCU's.

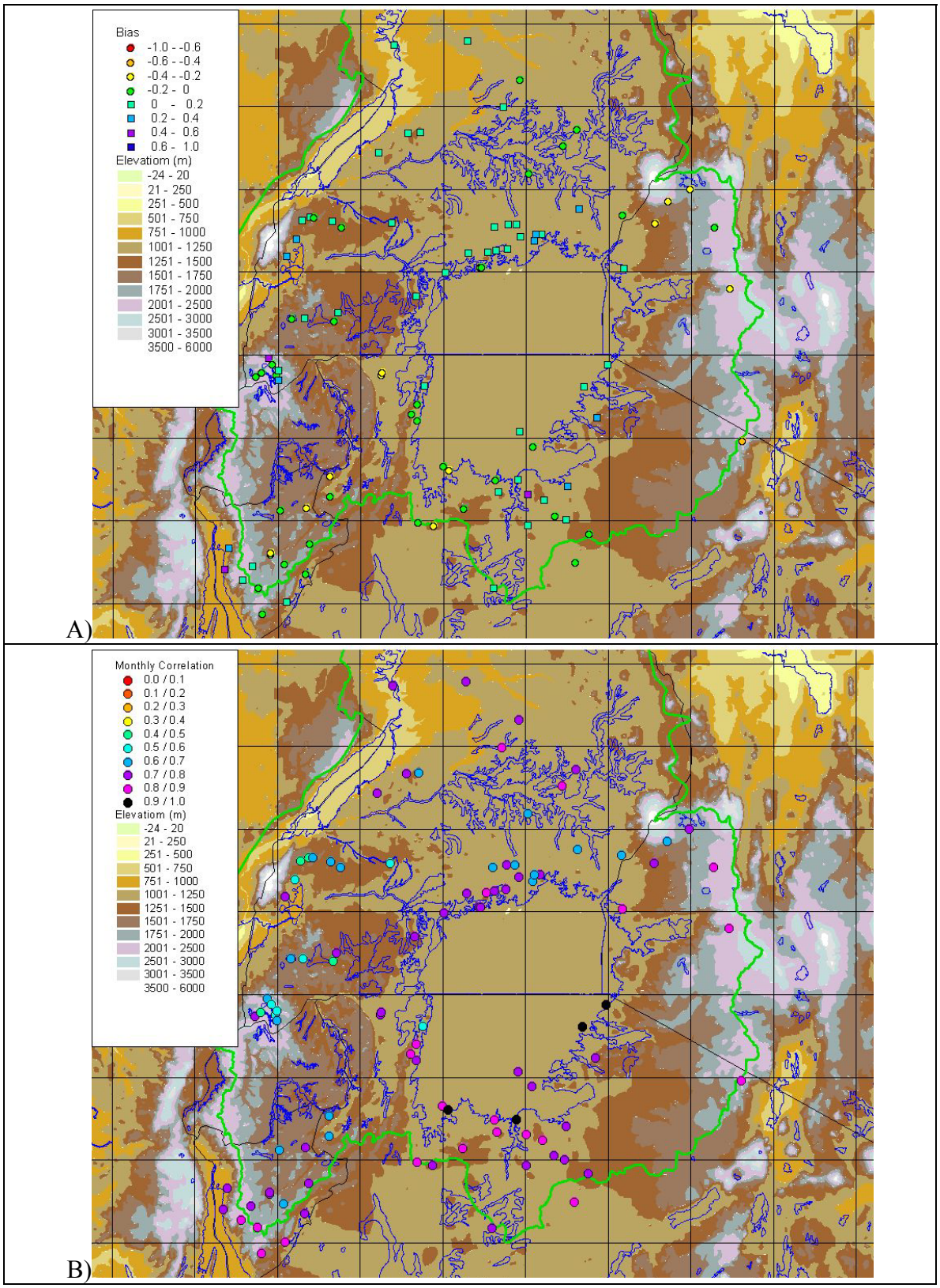


Figure 7.14 Relative bias and monthly correlation of the EIVWNCU estimator for the entire 1996-1997 dataset.

7.6 Summary

Two estimation procedures, one based on empirical precipitation distributions, and the other based on an analytical description of the precipitation probability density functions, were calibrated and validated over a three-year period using a dataset of more than 100 gages. The gages were distributed over a variety of climatic regions in the Lake Victoria basin, including lake islands, wet coastal areas on the northwestern lakeshore, dry coastal areas along the southeastern lakeshore, hilly areas, windward and leeward regions, and high mountains. The estimators showed good average performances and a good degree of robustness. When storm state and water vapor data were included, satellite-to gage correlation and bias improved. The largest improvement, however, was produced by the inclusion of VIS data. Consideration of seasonality improved the estimation of monthly precipitation, but at the expense of the estimation at the daily level, due to the limited amount of TRMM data. A critical factor for correctly reproducing the precipitation variability was the inclusion of precipitation temporal autocorrelation.

A larger database of coincident TRMM and geostationary data is critical to improve model performances at all resolutions.

CHAPTER 8

MULTI-PIXEL ANALYSIS

8.1 Chapter Scope

Most hydrological applications where remotely sensed precipitation would be useful pertain to areas larger than a single 5x5 km pixel, requiring the aggregation of precipitation over several pixels. However, correct characterization of the statistical properties of precipitation for multiple pixels requires generating a spatially correlated random precipitation field. More specifically, considering the spatial correlation of instantaneous precipitation (Section 8.3):

$$Corr(R(x_1, y_1, t), R(x_2, y_2, t)) \neq 0 \qquad for \quad (x_1, y_1) \neq (x_2, y_2) \qquad (8.1)$$

the scope of this chapter is to develop procedures that can be used operationally to generate realizations of the precipitation field that are representative both of the mean areal precipitation as well as the spatial precipitation variability. In this effort, use will be made of the correlation structure that can be derived from the TRMM images and of the single pixel probability densities associated with the observed IR, VIS/WV, and stage conditions.

The chapter is organized as follows: Section 8.2 describes the procedure used to assess the model's ability to represent the precipitation variability over sets of pixels of different size. Section 8.3 examines the spatial correlation of the instantaneous precipitation field and its variation in space and time. The multi-pixel performances of the spatially uncorrelated models presented in Chapter 7 are discussed again in Section 8.4 to provide a better understanding of the effects of ignoring the precipitation spatial

correlation and the improvements realized by the procedures described in the following sections. Section 8.5 examines past methodologies for creating spatially correlated random fields and their applicability in the context of this research. Section 8.6 describes the approaches taken in this study and assesses their performance. Finally, Section 8.7 summarizes the research findings and discusses possible extensions.

8.2 Performance Indicators

Ideally, the capability of a remote sensing methodology to represent the mean areal precipitation (MAP) and its variability over areas of different size should be assessed by comparing gage-derived and satellite-derived MAP over areas of increasing size, using indicators similar to the ones adopted for the single-pixel analysis (Section 7.2). Given the scarcity and unequal distribution of gages over the Lake Victoria area, however, such an approach is not feasible for the entire basin. It was instead decided to focus the multi-pixel analysis on a 1°x1° square around the cities of Entebbe, Kampala, and Jinja on the northern shore of Lake Victoria (Figure 8.1) during the period 1996-1997. This area, which will be referred to as *EKJ area* in the rest of the chapter, features the highest density of gages in the basin, thus allowing a better assessment of the aggregation of the estimated precipitation over several pixels.

Even in this area, however, the number of gages is insufficient for reliably assessing the MAP, especially at the daily temporal resolution, for which the number of gages with consistent data is limited to six and the precipitation spatial correlation is low. For this reason, the analysis compares instead the gage-measured precipitation and the satellite-estimated precipitation over all possible combinations of gages in the square. For each combination of gages, the daily/decadal/monthly gage-derived precipitation is computed by aggregating the daily/decadal/monthly precipitation measured at each single gage in the combination. On the other hand, the satellite-derived precipitation is

composed by a set of values, obtained by averaging satellite estimates belonging to the same realization of the precipitation random field for the pixels containing the gages.

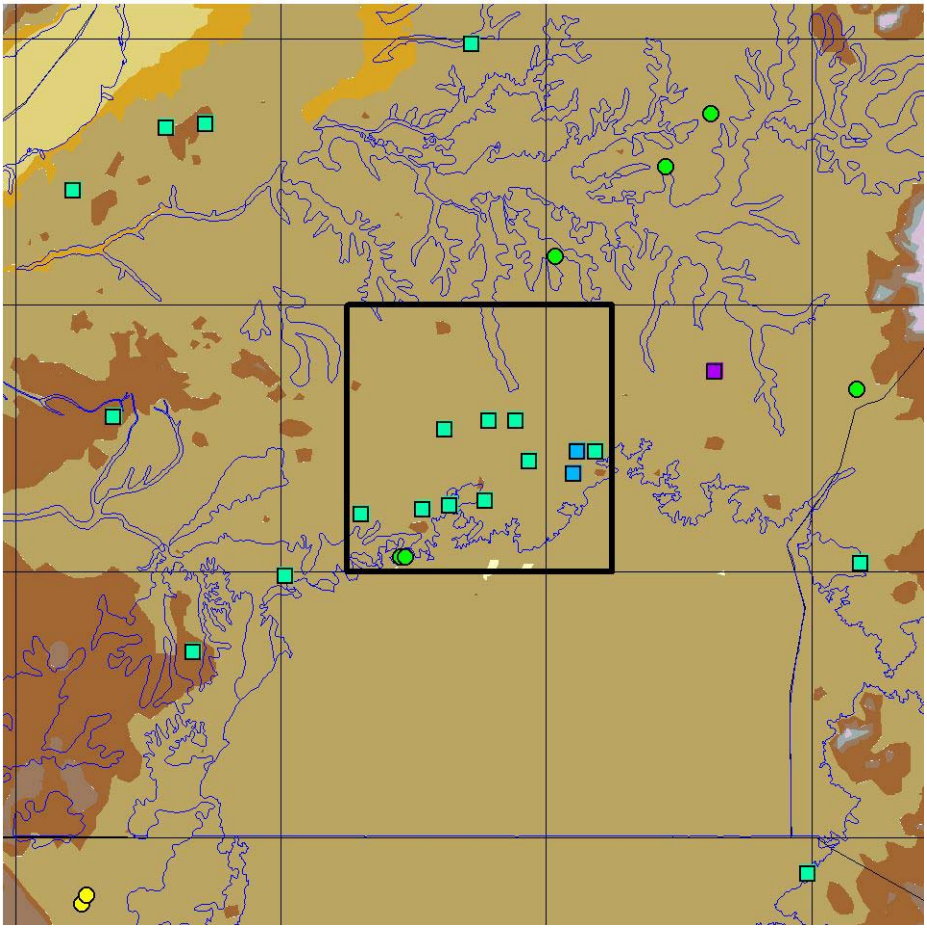


Figure 8.1 Location and gage distribution of the EKJ area (black square) in 1996-1997.

Each combination of gages is characterized by two parameters: the number of gages and the size *d* of the smallest square containing them. This size is an indication of how spread are the gages composing the combination. Furthermore, the average performance of all combinations of gages contained within a *dxd* square can be used to characterize the model ability to represent MAP over *dxd* spatial scales. For example, Figure 8.2 shows that the satellite-gage correlation is increasing with the number of gages

considered. The increase in the correlation is not linear, but more of a logarithmic type. Furthermore, satellite-gage correlation increases less when the pixels are close to each other than when they are widely separated. These two observations lead to argue that the satellite-gage correlation of the MAP is probably around 0.65 for areas smaller than 5x5 pixels (i.e., 0.25°x0.25°) and above 0.7 for areas larger than 10x10 pixels (i.e., 0.5°x0.5°).

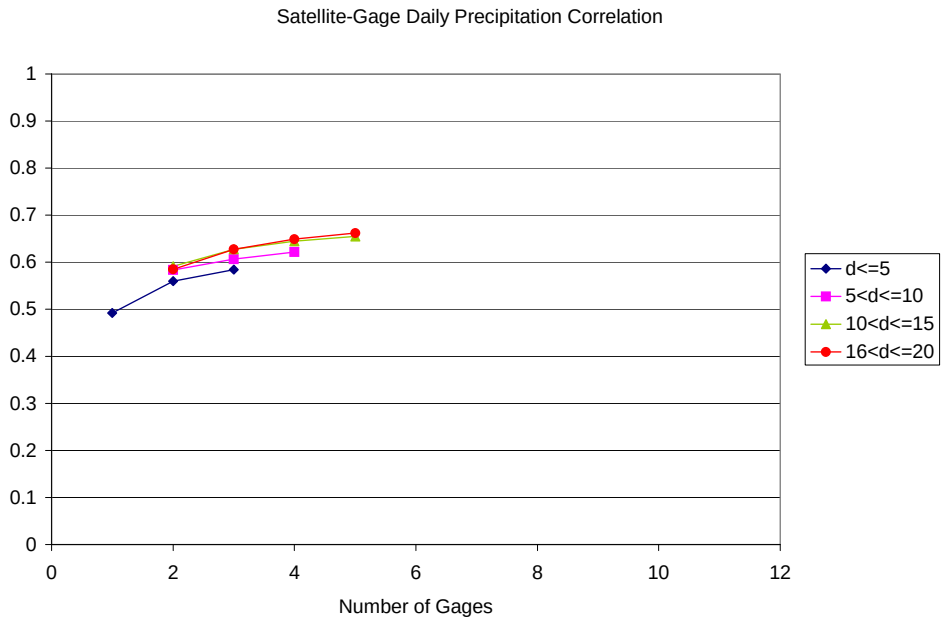


Figure 8.2 Satellite-gage correlation of daily precipitation as a function of the number of gages and size d of the smallest square containing the combination of gages.

The previous observation shows that this two dimensional relationship (number of gages and minimum distance) can only be captured by charts such as the above. However, to avoid the inclusion of an overwhelming number of charts, only three measures will be presented here: (1) the correlation between gage data and the average of the ensemble, (2) the mean absolute error between gage data and the average of the

ensemble, and (3) the 95% compliance rate. The first two indicators assess the ability of the ensemble average to track the behavior of the gage records, while the third assesses the ability of the ensemble to represent the uncertainty of the precipitation patterns in a statistically meaningful way. Furthermore, the complete analysis will be reserved for the final models, while a simplified analysis will be presented in the preliminary steps.

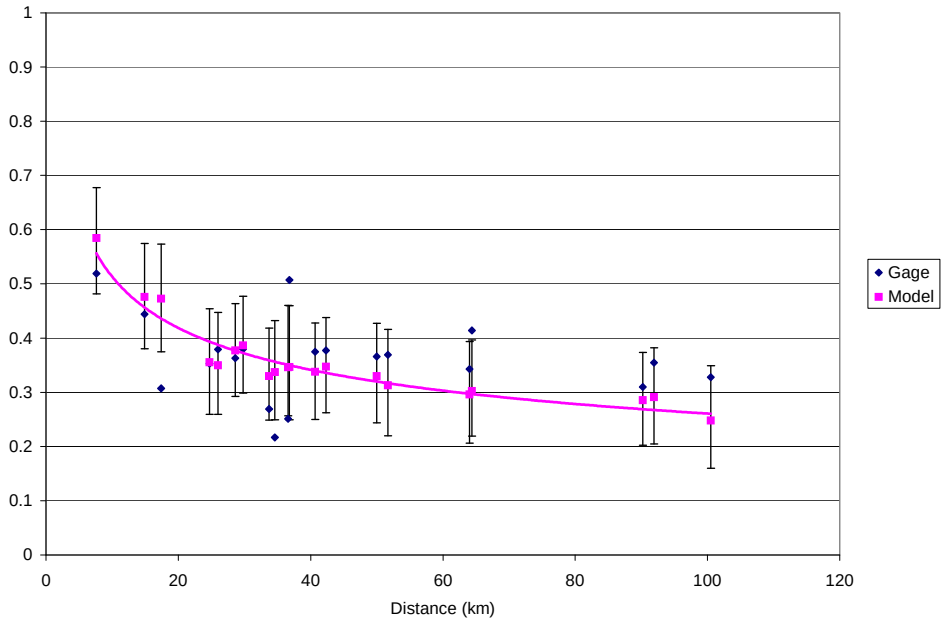


Figure 8.3 Correlation in daily precipitation in the EKJ area in 1996-1997. The blue marks indicate the correlation of gage-measured precipitation. Model results represent the correlation of each realization of the satellite-estimated precipitation. The purple marks correspond to the average of the correlation spectrum and the interval bars to its 2.5% and 97.5% bounds.

Another element of interest is the model ability to reproduce the spatial correlation of precipitation. This is evaluated at two temporal levels: at the half-hour level, the precipitation's spatial correlation of a number of representative TRMM passages is compared with the corresponding spatial correlation of each member of the estimation ensemble. At the daily level, the correlation in the precipitation measured at each pair of gages during the entire period 1996-1997 is compared with the correlation of

the precipitation in the pixels containing the gages produced by each individual member of the estimation ensemble. Showing the correlation curves of each of the 500 members of the estimate ensemble would not be very effective. Thus, for each distance only the average, the 2.5% lower bound, and the 97.5% upper bound of the correlation spectrum are shown (Figure 8.3).

8.3 Precipitation Spatial Correlation

Daily precipitation is strongly spatially correlated. This is because individual storms generally affect contiguous areas of several hundreds of square kilometers. Furthermore, the overall storm precipitation tends to be concentrated around core areas and tapers off towards the storm edges. Over longer periods, average precipitation is more strongly correlated, because, due to averaging, precipitation becomes a function of the climate of the region, which generally varies smoothly over large areas. Exceptions to these general behaviors are normally associated with significant geographic features, such as mountain ridges, narrow valleys, and sea/land interfaces, which create abrupt changes in precipitation patterns.

What may be less intuitive is that even instantaneous precipitation exhibits a significant spatial correlation, especially over short distances. For example, Figure 8.4 shows that the instantaneous precipitation over the Lake Victoria basin is considerably correlated even up to 25-30 km.

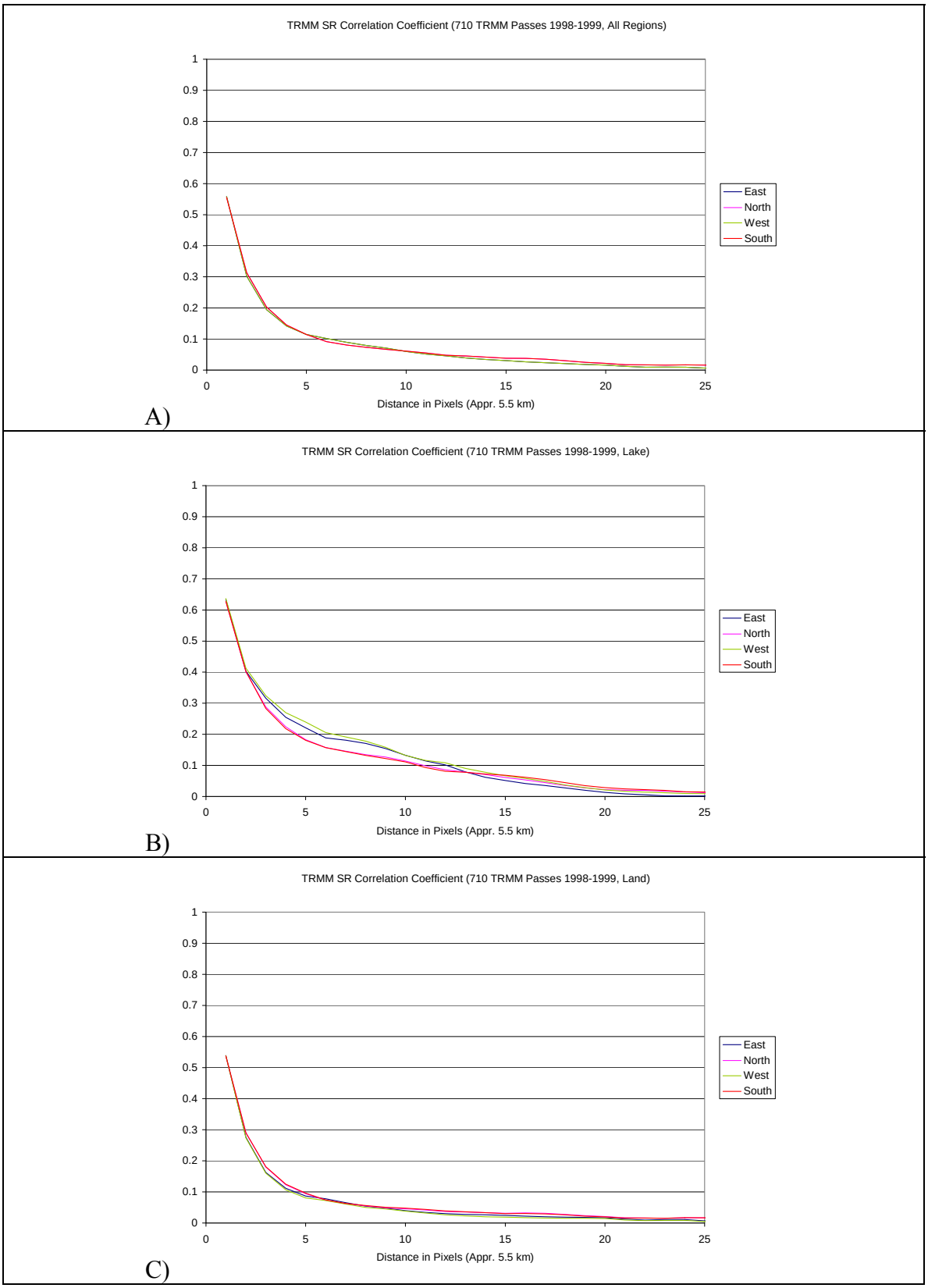


Figure 8.4 Instantaneous precipitation spatial correlation in the Lake Victoria region as a function of distance in pixels and direction. A) Entire basin; B) Lake area; C) Land area;

The spatial correlation is higher over the lake than over the surrounding watershed, due to the larger and more organized storms affecting the lake and to the strong orographic features present in the watershed. Interestingly, the precipitation correlation over the lake is slightly stronger along the East-West axis than along the North-South axis, reflecting the westward direction of winds and propagation of most storms. Over the lake watershed, local smaller storms, orographic rain, and leeward areas mask this effect. Also, the average precipitation correlation over the lake is very similar to that found by Bell et al. (1990) for the tropical Atlantic during the GATE experiment, showing that the large extension of the lake induces a certain maritime precipitation behavior, very different from the surrounding watershed.

As shown in Figure 8.5, small storms are characterized by spatial correlation lower than the mean correlation, while large storms feature a spatial correlation that is much stronger and more persistent than the mean correlation, likely due to the large areas of persistent stratiform rain they create in their wake.

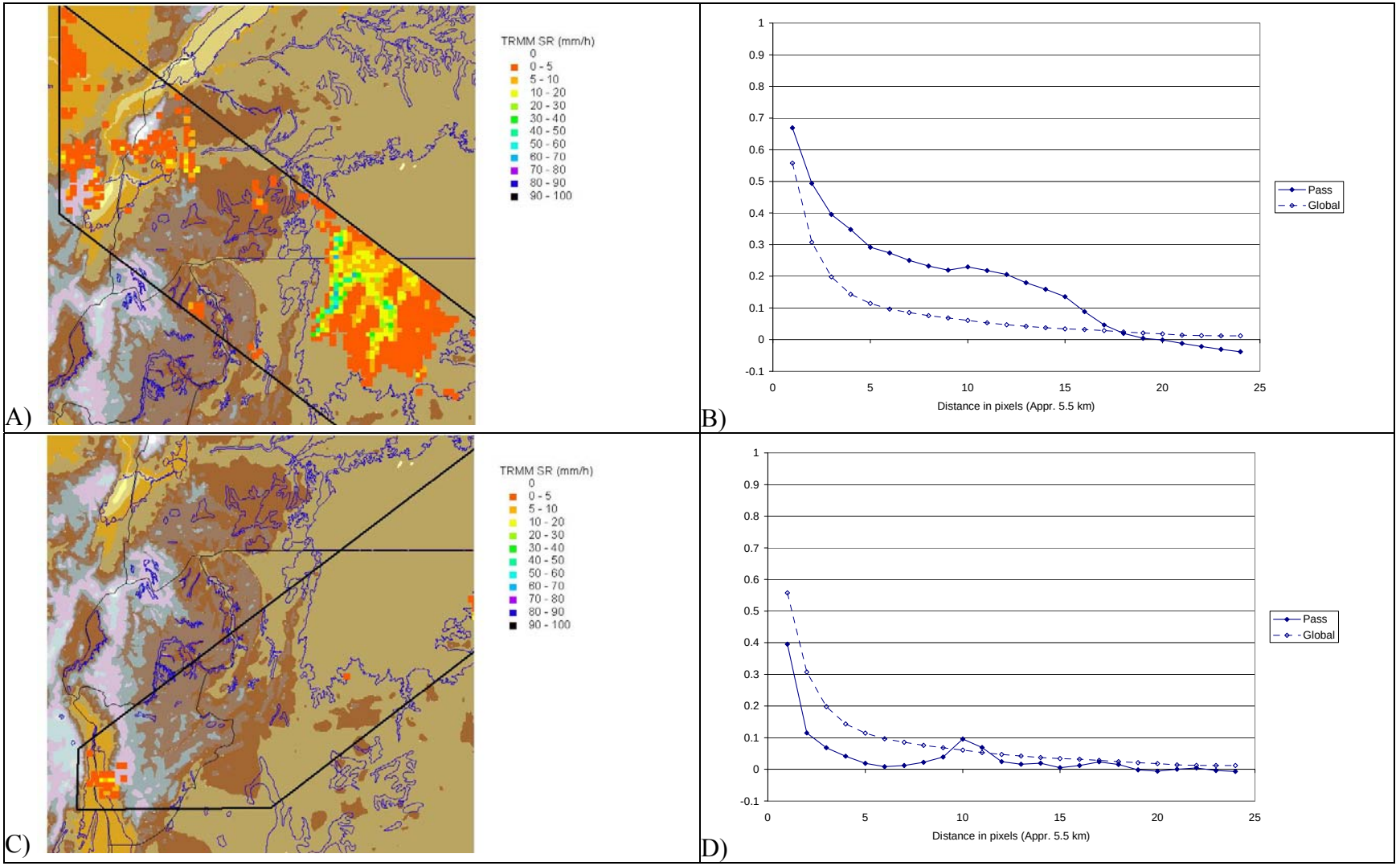


Figure 8.5 Spatial distribution and spatial correlation of TRMM PR Surface rain during single passes. A) and B) May 3, 1998 23:26 GMT; C) and D) May 4, 1998 10:50 GMT. The black solid line delimitates the PR swat.

8.4 Spatially Uncorrelated Models

Incorporating spatial correlation in the generation of the precipitation field is a computationally demanding task. Thus, the necessity of doing so will first be investigated by assessing the performance of spatially uncorrelated fields. Furthermore, the performance of the uncorrelated fields will provide a baseline against which to compare the performance of the spatially correlated fields.

Figure 8.6 shows the spatial distribution of one of the 500 realizations produced by the spatially and temporally uncorrelated empirical distribution model. These realizations correspond to the same TRMM over flights shown in Figure 8.5 as produced by the spatially and temporally uncorrelated empirical distribution model. The location of the observed precipitation is well replicated, as is, in part, the intensity of the precipitation. Yet, the spatial distribution of the simulated rain has no spatial continuity, with pixels featuring high rain randomly alternating with no-rain pixels. The resulting spatial correlation is much lower that the one observed in large storms. It is also lower than that observed in small storms over short distances.

One question that might be asked is whether integrating the half-hour precipitation to produce the daily/dekadal/monthly precipitation could sufficiently smooth the rain field to create realizations that are more credible. This is investigated in Figure 8.7, which shows that the smoothed spatially uncorrelated precipitation matches the average gage spatial correlation only for distances larger than 20 km. It also shows that the analytical distribution model is noisier than the empirical distribution model. The failure to replicate the precipitation spatial correlation does not affect the ability of the ensemble average to estimate observed precipitation, especially for large areas (Figure 8.8):

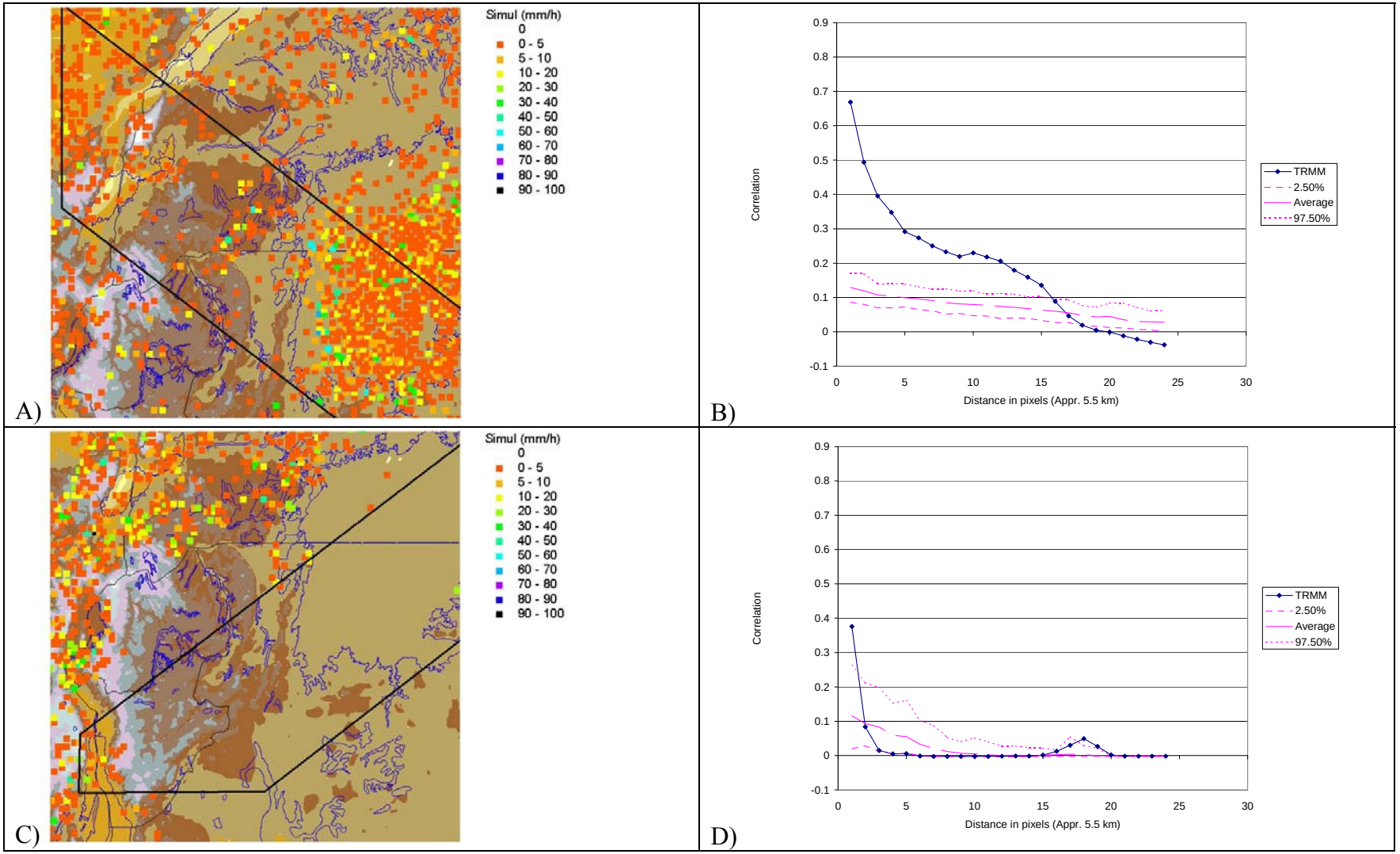


Figure 8.6 Realizations of simulated precipitation and spatial autocorrelation of observed (solid) and simulated (dashed) precipitation using spatially and temporally uncorrelated empirical distributions. A) and B) May 3, 1998 23:26 GMT; C) and D) May 4, 1998 10:50 GMT.

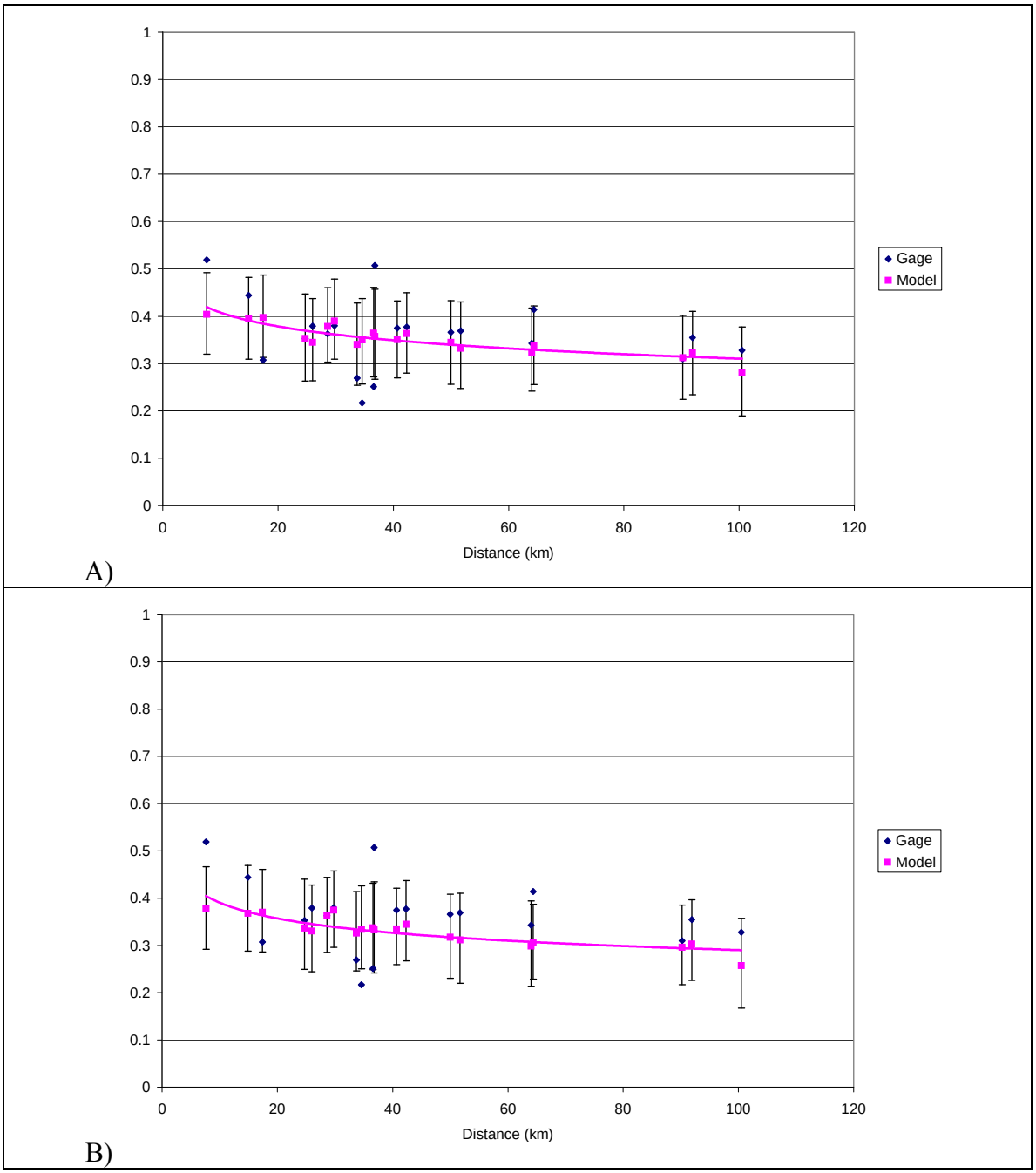


Figure 8.7 Correlation in daily precipitation for the EKJ area in 1996-1997, temporally and spatially uncorrelated models. A) Empirical distributions. B) Analytical distributions.

- At the monthly level, the gage-satellite correlation increases from 0.76 for the single pixel to 0.90 for combinations of nine gages, while the MAE decreases from 37% to 22% of the average precipitation.

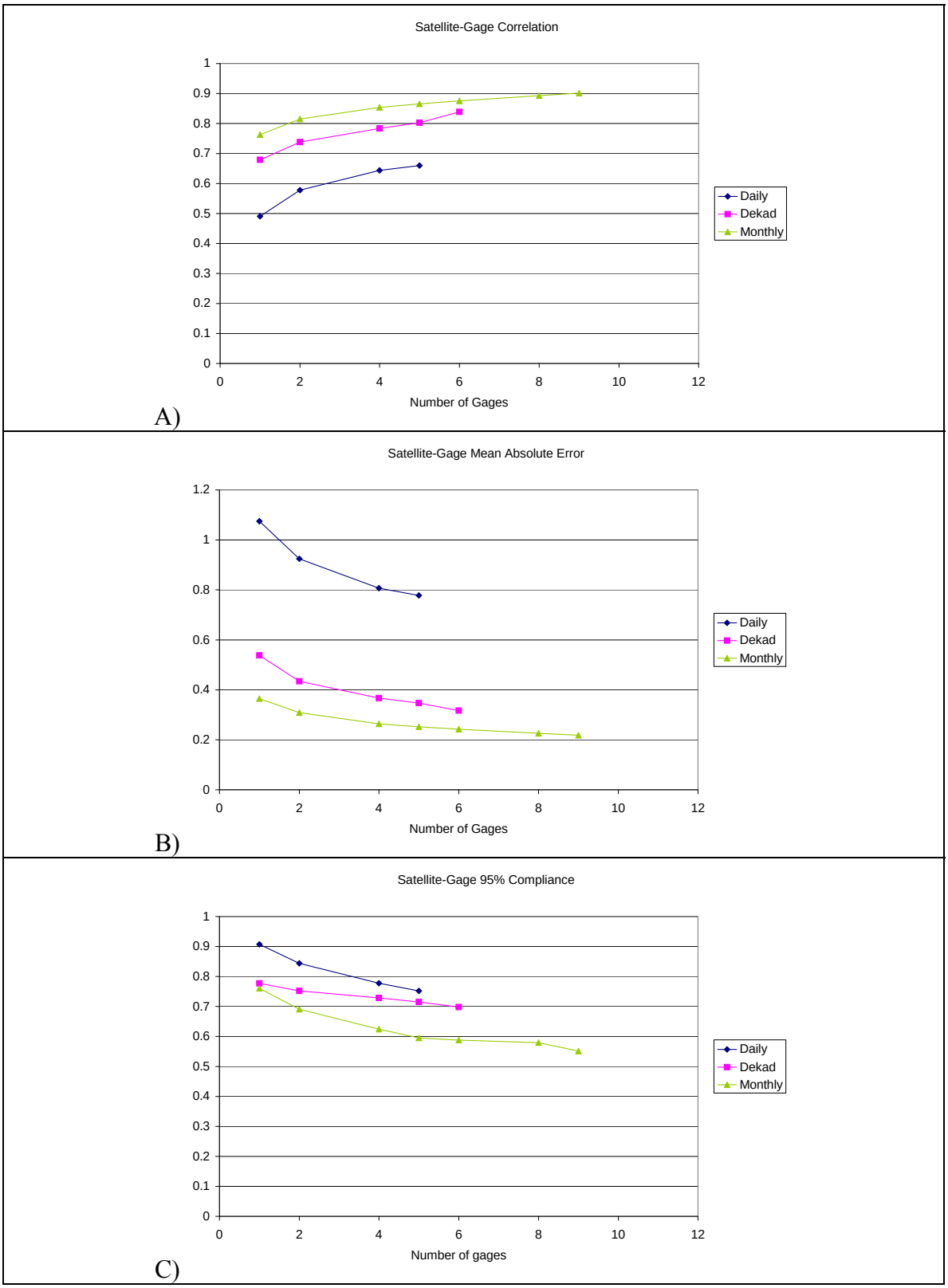


Figure 8.8 Multipixel performances for temporally and spatially uncorrelated empirical distribution simulation in the EKJ area.

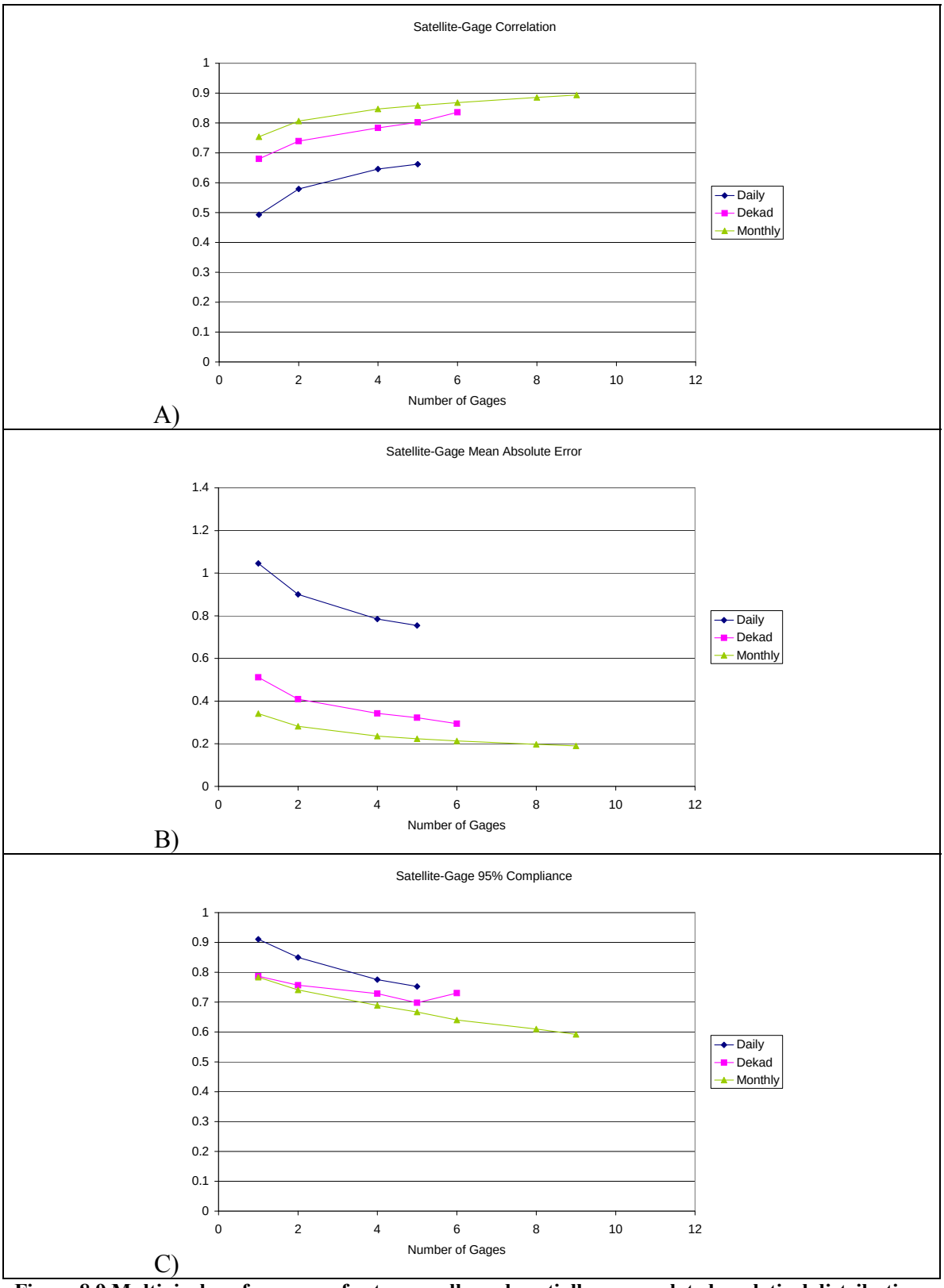


Figure 8.9 Multipixel performances for temporally and spatially uncorrelated analytical distribution simulation in the EKJ area.

- At the daily level, the gage-satellite correlation increases from 0.49 for the single pixel to 0.66 for combinations of five gages, while the MAE decreases from 108% to 78% of the average precipitation.

On the other hand, the 95% compliance rate decreases more than 20 percentage points at the monthly level and more than 16 percentage points at the daily level, showing that the spatially uncorrelated models are not able to replicate the precipitation variability.

The analytical distributions behave similarly, with just a slightly better MAE and 95% compliance rate, probably due to the lower bias of this model.

The temporally correlated distributions have a spatial correlation that is more than 15 percentage points lower than their corresponding temporally uncorrelated versions and the gage data (Figure 8.10). This occurs because, under the temporally correlated model, the precipitation of a pixel at time *t* depends only on the precipitation of the same pixel at time *t-1*. The temporal correlation amplifies the effect of the lack of spatial correlation, since markedly different precipitation rates at the beginning of a storm may result in completely different precipitation totals even in adjacent pixels. Interestingly, two completely different ways to account for the temporal correlation (empirical and analytical distributions) have a very similar effect, with the analytical distribution having only a slightly lower spatial correlation, exactly as in the temporally-uncorrelated case.

The increase in the correlation between gage data and the average of the estimation ensemble is similar to the temporally uncorrelated case (Figure 8.11 and Figure 8.12), but because of the lower single pixel satellite-to-gage correlation the empirical distributions achieve a lower correlation than the temporally-uncorrelated model and the temporally-correlated analytical distributions.

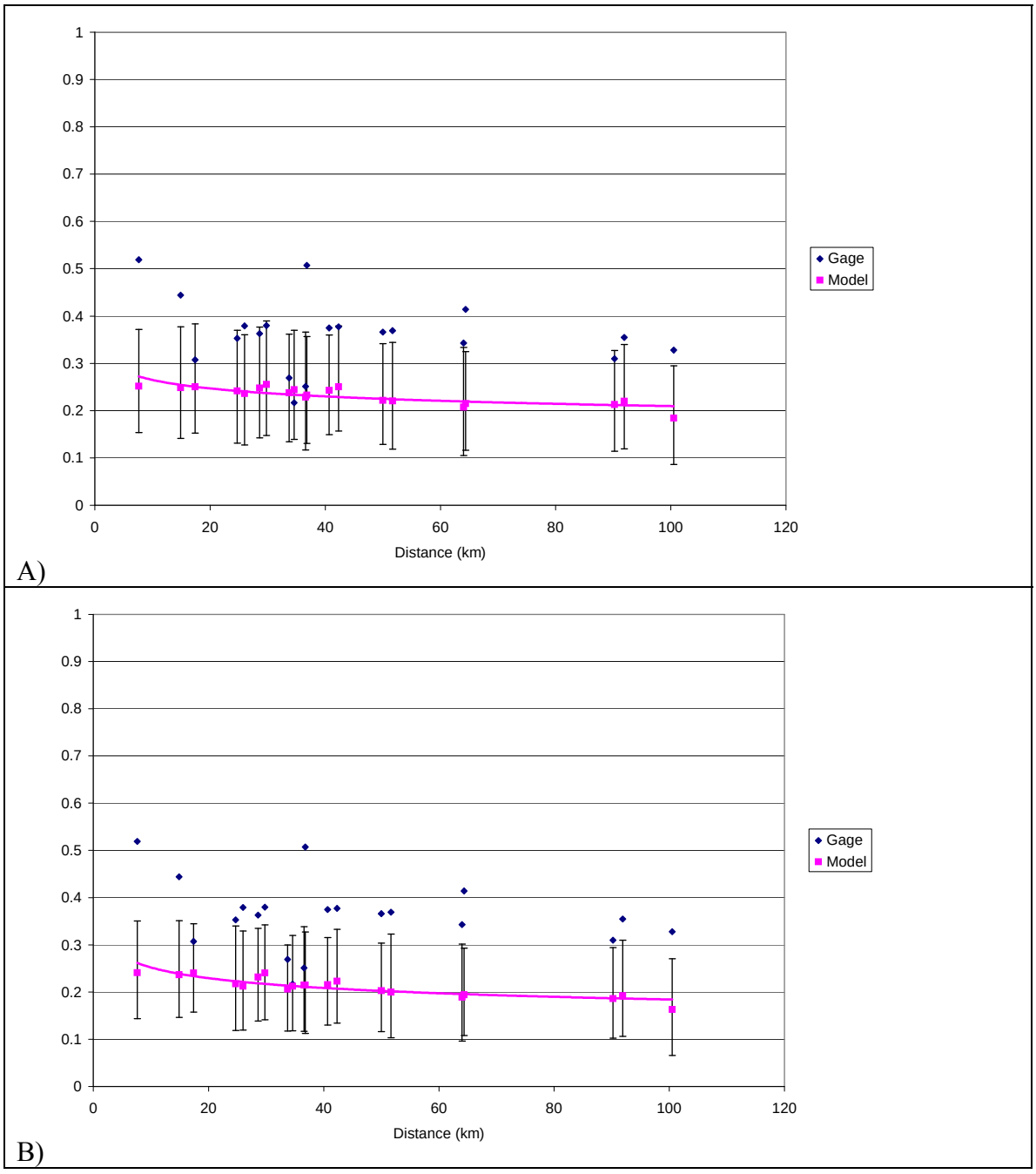


Figure 8.10 Correlation in daily precipitation for the EKJ area in 1996-1997, temporally-correlated/spatially-uncorrelated models. A) Empirical distributions. B) Analytical distributions.

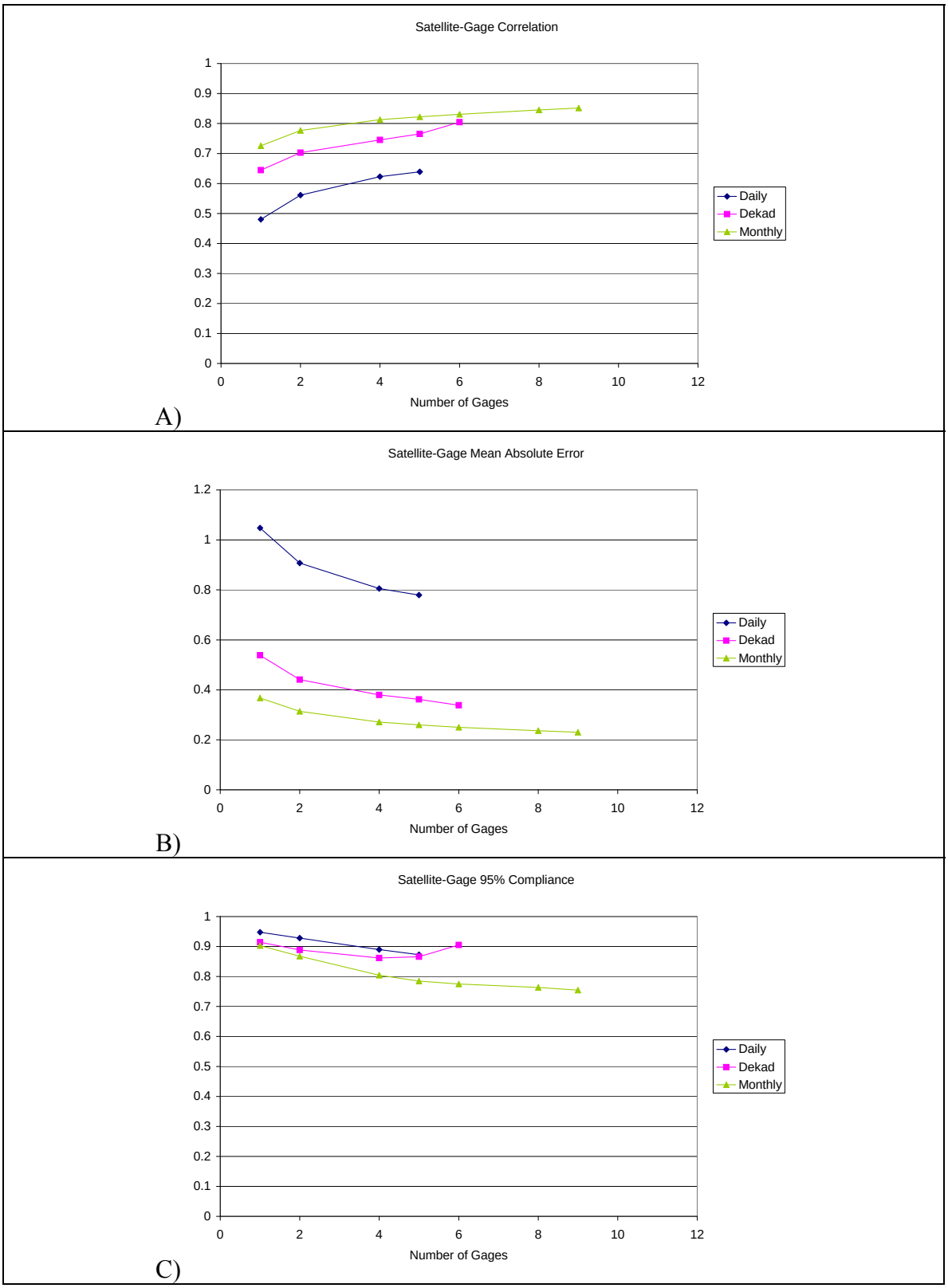


Figure 8.11 Multipixel performances for temporally-correlated/spatially-uncorrelated empirical distribution simulation in the EKJ area.

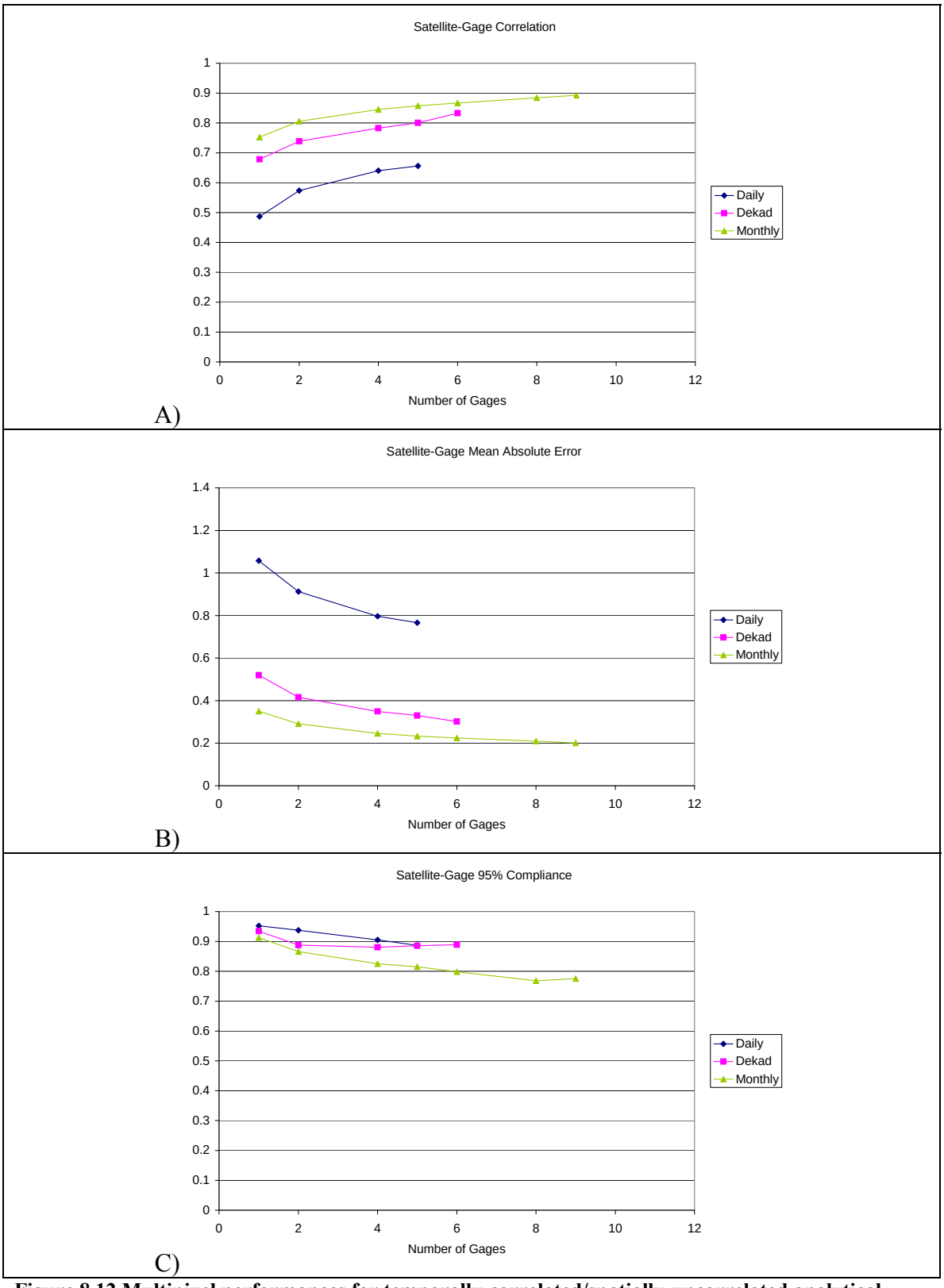


Figure 8.12 Multipixel performances for temporally-correlated/spatially-uncorrelated analytical distribution simulation in the EKJ area.

The latter model also shows a lower MAE. On the other hand, the higher single-pixel 95% compliance rate increases the overall 95% compliance rate of the multi-pixel combinations. Furthermore, the downward decline caused by increasing the number of the gages is reduced: for example, the 95% compliance of the daily precipitation in the analytical distributions decreases by 16 percentage points for the temporally-uncorrelated version, but by only 6 percentage points for the temporally correlated model.

In summary, the ability of the uncorrelated random fields to represent the precipitation variability over large areas significantly decreases with the size of the area. The inclusion of spatial precipitation correlation is, therefore, necessary when one is interested to represent the precipitation variability over large areas.

8.5 Generation of Spatially Correlated Random Fields

As illustrated in Chapter 7, the generation of realizations of a spatially uncorrelated (or orthogonal) random field is done simply by sampling the empirical or analytical distribution of precipitation pixel by pixel. In this way, the ensemble of possible realizations will reproduce for each pixel the precipitation distribution associated with the observed satellite data and geography. This distribution in the rest of the chapter will be referred to as *prior distribution*, *unconditional distribution*, or *satellite distribution*.

Reproducing a spatially correlated random field is a more difficult task because the ensemble of realizations must converge not only to the desired mean and variance, but also to the desired correlation structure. Several methods have been proposed to accomplish this goal, depending on the characteristics of the random field to be reproduced. The following section briefly reviews some of the methodologies proposed and analyzes their ability to generate a precipitation field utilizing the information provided by satellites.

8.5.1 Correlated Random Field Generation by Spectral Methods

Spectral methods take advantage of the fact that n-dimensional random fields in the space-time domain have a dual representation as n-dimensional random fields in the frequency domain. More specifically, each random field can be represented as the superposition of a large (even infinite) number of sinusoidal waves of a certain frequency and amplitude given by a random variable. The set of these random variables is called "*spectral representation*" of the spatial random field and is related to it through the Fourier transform:

$$Z(\mathbf{x}) = \sum_{\mathbf{k}} a_{\mathbf{k}} e^{i\mathbf{k}^T\mathbf{x}} \tag{8.2}$$

where **x** and **k** are n=dimensional vectors.

If the random field *Z* is real, multivariate Gaussian, stationary, and zero mean, then its spectral representation is a spatially uncorrelated random field, which is very easy to compute. In particular, the real and imaginary parts of the coefficients ***a(k)*** are uncorrelated Gaussian random variables excepts for the constraints required to generate a real random field. The variances of these random coefficients depend on *k* and are determined by the desired spatial correlation for the field ***g(x)*** according to:

$$\left\langle a_k^* a_k \right\rangle = N^{-2} \sum_x c_g(x) e^{-i\mathbf{k}^T\mathbf{x}} \tag{8.3}$$

where $c_g(\mathbf{x})$ is the correlation of g at two grid points separated by **x**, and * is the conjugate operator (Bell, 1987).

A two-dimensional Gaussian field ***g(x)*** with zero mean and unit variance over a NxN Cartesian grid can be generated by the following scheme (Gremont, 2002):

1. Generate 2D uncorrelated complex noise, $e(k_1,k_2)$, with real and imaginary parts with zero means and unit variances, under the constraint $e(k_1,k_2)=(e(n_1N-k_1, n_2N-k_2)^*$, $n_1, n_2 \in \mathbb{N}$.
2. Create a matrix $h(k_1, k_2) = ½ \; \forall(k_1,k_2)$, $h(0,0)=1$, $h(0,N/2)=1$, $h(N/2,0)=1$, $h(N/2,N/2)=1$
3. Compute the $C_g^f(k_1, k_2)$, the 2D-FFT of $C_g(X_1, X_2)$
4. Compute the complex matrix A whose elements are obtained as: $a(k_1,k_2)=h(k_1,k_2)* C_g^f(k_1, k_2),*e(k_1,k_2)$
5. Obtain the spatially correlated field by computing the 2D inverse fast Fourier transform (FFT) of the matrix A.

If the original variables can be transformed into a correlated stationary Gaussian field, then this procedure can generate it. In this case, the imposed correlation is the correlation of the transformed variable. Bell (1987) describes a procedure to transform a normal field into a precipitation field that has intensity 0 with probability P_0 and lognormally distributed intensities. A key of this procedure, however, is that there must be a unique relation between the Gaussian field **g** and the rain field, so that it is possible to determine the correlation structure of g that will produce the precipitation correlation after the transformation. This is possible only if a unique set of parameters (P_0, μ, σ) is used for the precipitation model.

8.5.2 Correlated Random Field Generation by Turning Bands Methods

First introduced by Matheron in 1973 and further improved by Mantoglou and Wilson (1982) and others, this method reduces the generation of 2D and 3D, spatially correlated, realizations of a random field to the generation of one-dimensional random processes. The one-dimensional correlation structure along each line is uniquely derived

from the original 2D or 3D correlation structure by a deconvolution process (Mantoglou and Wilson, 1982). A 1D random process is generated for each line via a 1D spectral method similar to the one described in the previous section (Mantoglou and Wilson, 1982).

The algorithm proceeds to select a "*star*" composed of *L* lines with common midpoint and uniform distribution on the unit circle or sphere. For each line, it then generates the corresponding 1D autocorrelated process. The minimum number *L* of lines is between 4 and 16 for 2D processes and 15 for 3D processes (Mantoglou and Wilson, 1982). The value of the 2D or 3D random field at a point **x** is computed by projecting the point onto each of the **L** lines and combining the corresponding values generated by the 1D-processes (Mantoglou and Wilson, 1982). More specifically,

$$z(\mathbf{X}) = \frac{1}{\sqrt{L}} \sum_{i=1}^{L} z_i(\mathbf{X} \cdot \mathbf{u}_i) \tag{8.4}$$

where: z_i = 1D line process realizations;

$\mathbf{u}_i$ = direction vector of line i.

The 2D or 3D process must be second-order stationary, isotropic, normally distributed, and zero mean (Mantoglou and Wilson, 1982).

The turning bands method is faster than the FFT based methods, especially for a large number of points. However, the realizations show that the computational process tends to introduce artificial behavior in 3D due to the limited number of lines (15) used for partitioning the space. Deutsch and Journel, 1998 also note difficulties in handling anisotropic processes,

8.5.3 Correlated Random Field Generation by Matrix Decomposition

The matrix decomposition method (Alabert, 1987; Davis, 1987) provides an alternative, and in certain cases very effective, way to simulate spatially correlated Gaussian fields. The covariance between each pair of the N points to be simulated is computed and stored as a symmetric matrix $\boldsymbol{C}$ of size $N^2 x N^2$. Although the covariance of the Gaussian process does not need to be stationary, it is assumed that the mean of the distribution is zero everywhere. The matrix $\boldsymbol{C}$ is then decomposed into the lower and upper triangular matrices $\boldsymbol{L}$ and $\boldsymbol{U}$.

$$\mathbf{C}=\mathbf{LU} \qquad \mathbf{L}=\mathbf{U}^{T} \tag{8.5}$$

The desired random field $\boldsymbol{Z}$ is obtained multiplying $\boldsymbol{L}$ by a vector $\boldsymbol{b}$ of N uncorrelated, normally distributed random variables.

$$\mathbf{Z}=\mathbf{Lb} \tag{8.6}$$

This method is very effective for cases where many realizations of the same random field must be produced, since, after the matrix $\boldsymbol{C}$ is decomposed, each realization of $\boldsymbol{Z}$ is quickly obtained from equation 8.6 by generating an array of N uncorrelated, normally distributed. On the other hand, the size of $\boldsymbol{C}$ grows rapidly, limiting the value of N to few hundreds (Deutsch and Journel, 1998). Attempts to partition a larger domain into smaller overlapping subsets that would allow the use of the matrix partition method have not been completely successful.

8.5.4 Correlated Random Field Generation by Sequential Simulation

Sequential simulation is probably the most popular stochastic algorithm for producing realizations of random fields in Earth sciences, thanks to its computational

efficiency and capacity of replicating the variogram and other characteristics of the data distribution. It was introduced by Journel (1989) for generating stationary random fields with non-parametric probability distribution (sequential indicator simulation, or SISIM). The sequential simulation concept was later adapted for generating Gaussian multivariate random fields by Gomez-Hernandez, 1991 (sequential Gaussian simulation, or SGSIM). More recently, several authors have tried to apply the concepts of the less computationally intensive SGSIM to the simulation of non-Gaussian distributions through the so-called direct sequential simulation (DSSIM).

The premise of all the sequential simulation algorithms is that for any set $\{Z(x_j), j=1,\ldots,N\}$ of random variables defined at N locations x_j, and M known data $\{z(x_k), k=1, \ldots, M\}$, the joint multivariate conditional cdf can be decomposed as (Caers, 2000):

$$F(x_1, \ldots, x_N; z_1, \ldots, z_N \mid (M)) = F(x_N; z_N \mid (M+N-1)) \cdot \ldots \cdot F(x_2; z_2 \mid (M+1)) \cdot F(x_1; z_1 \mid (M)) \quad (8.7)$$

Where $F(x_k; z_k \mid (M+k-1))$ is the conditional cdf of $Z(x_k)$ given the set of the M "known" data and the previous $k-1$ realizations $\{z^l(x_j), j=1,\ldots,k-1\}$.

This relation allows the generation of a random field by sequentially inspecting all its nodes according to the following algorithm (Caers, 2000):

1. Perform a variable transformation if needed;
2. Define a random path visiting all nodes;
3. For each node x_j in the random path sequence;
    a. Model the conditional distribution $F(x_j; z_j \mid (M+j-1))$ given the known data and the previously simulated $j-1$ nodes;
    b. Draw the simulated value x_j from $F(x_j; z_j \mid (M+j-1))$;
4. If needed, perform a back-transformation to reproduce the target histogram.

The sequential simulation approach can be applied to very different random fields and very different conditioning information by just changing the way steps 3.a and 3.b are implemented.

One potential drawback of the sequential simulation is that a new set of equations must be generated every time a new node is simulated and added to the set of conditioning points. If the number of points is high this process can take a long time, especially if the random path used for visiting the nodes is changed at every simulation (e.g., Deutsch and Journel, 1998). For this reason, most algorithms use only a limited number of the closest measured and previously simulated data to condition the generation of a new value. The idea is that the closest data mask the influence of farther nodes, which have a lower correlation (Emery, 2004). A problem with this approach is that the statistical properties of the known data and previously simulated data are reproduced only up to the maximum distance present in the limited set of conditioning values. This is particularly true in the latest stages of the simulation, when the density of previously simulated values is high, screening off farther values. A way to minimize this effect is by adopting a multi-grid search strategy in which the simulation proceeds along a series of grids of decreasing coarseness (for example, nodes are randomly chosen along a grid of resolution 16x16 nodes first, then along a grid of resolution 12x12, etc.). The nodes of the coarsest grid are relatively sparse, allowing the reproduction of the large-scale characteristics of the covariance.

8.5.4.1 A Note About Kriging

Although other approaches are possible, step 3.a of the sequential simulation algorithm is normally accomplished using some type of a kriging scheme. As mentioned in Section 2.2.1, the problem of estimating the value of a variable Z at a location x_0 given the values the variable assumes in M nearby locations can be treated in various ways.

The most common way to deal with this problem is to express *z(x0)* as a linear combination of the *M* known values, namely,

$$z(x_0) = a_0 + \sum_{k=1}^{M} a_k z(x_k) \tag{8.8}$$

The key of the approach is in the computation of the coefficients a_k. One possibility is to compute them as a function of just the geographical distance between the target point and the known values (Thiessen polygons, inverse (square) distance methods, etc.). If historical records are available for x_0 and the neighboring points, then the coefficients a_k can be chosen to minimize the mean square error of the estimate. This approach, however, can practically be followed only for a limited number of locations in a large area.

Kriging procedures, on the other hand, assume $Z(x_0)$ as an element of a random field $Z(x)$ with given mean $m(x)$ and covariance $C(x_k,x_j)$, and the measured values $z(x_k)$, $k=1, \ldots,M$ as one particular realization of the random field. This means, that 8.8 can be seen as just a realization of the more general estimator:

$$Z^*(x_0) = a_0 + \sum_{k=1}^{M} a_k Z(x_k) \tag{8.9}$$

The coefficients a_k are then determined by imposing certain conditions on the estimation error $Z(x_0)-Z^*(x_0)$. The first condition is that the expectation of the estimation error be zero, namely (Journel, 1989):

$$\begin{aligned} &E(Z(x_0) - Z^*(x_0)) = E(Z(x_0)) - E(Z^*(x_0)) = E\{Z(x_0)\} - a_0 - \sum_{k=1}^{M} a_k m(x_k) = 0 \\ &\Rightarrow a_0 = m(x_0) - \sum_{k=1}^{M} a_k m(x_k) \end{aligned} \tag{8.10}$$

Thus:

$$Z^*(x_0) = m(x_0) + \sum_{k=1}^{M} a_k (Z(x_k) - m(x_k)) \qquad (8.11)$$

Equation 8.11 shows that the optimal estimator is obtained by adding to the a-priori knowledge of the field ($m(x_0)$) a linear combination of the differences between the measured values (i.e., the realizations of the random variable $Z(x_k)$) and their a-priori estimates ($m(x_k)$).

The second condition is that the variance of the estimation error be minimal, which results in the set of equations known as "*normal system*" (Journel, 1989):

$$\sum_{l=1}^{M} a_l C(x_l, x_j) = C(x_0, x_j) \qquad \text{for } j = 1,...,M \qquad (8.12)$$

The estimation variance is given by:

$$VAR\{Z(x_0) - Z^*(x_0)\} = C(x_0, x_0) - \sum_{k=1}^{M} a_k C(x_0, x_k) \qquad (8.13)$$

Equation 8.12 illustrates that the variance of $Z^*(x_0)$ does not depend on the values of random variables in the surrounding pixels, but only on the covariance structure and on the relative location of the known and previously simulated values. Equation 8.12 can be rewritten in matrix notation as:

Ka=k ,

$$\mathbf{K} = \begin{bmatrix} C(x_1,x_1) & C(x_1,x_2) & \dots & C(x_1,x_n) \\ C(x_2,x_1) & C(x_2,x_2) & \dots & C(x_2,x_1) \\ \dots & \dots & \dots & \dots \\ C(x_n,x_1) & C(x_n,x_1) & & C(x_n,x_n) \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix}, \quad \mathbf{k} = \begin{bmatrix} C(x_1,x_o) \\ C(x_2,x_0) \\ \\ C(x_n,x_o) \end{bmatrix} \tag{8.14}$$

The matrix ***K*** is real and symmetric, but to ensure that 8.14 has a unique solution, it must also be positive definite, imposing a constraint on the shape of the covariance function $C(x_k,x_j)$. Equations 8.11, 8.12, and 8.13 define a *simple kriging* estimator. Other forms of kriging estimators may be used according to the information available on the random field *Z*.

8.5.4.2 Sequential Gaussian Simulation

If the underlying multivariate distribution is assumed to be normal and stationary, then all conditional distributions of any subset of the random space $Y(\mathbf{x})$ given realizations of any other subset are also multivariate normal. What is particularly important for the sequential simulation is that any univariate conditional distribution given any other subset is univariate normal. Consequently, it is completely identified by its mean and variance, which can be determined from the known and previously simulated data through kriging. For, example if the mean and covariance of the distribution is known at any point of the random fields, as shown previously, the simple kriging can be used to estimate these two parameters (Deutsch and Journel, 1998):

$$\begin{aligned} E\{Z(x_j) \mid z(x_k), K = 1,..,M+j-1\} &= [z(x_j)]_{SK} \\ &= m(x_j) + \sum_{k=1}^{M+j-1} a_k [z(x_k) - m(x_k)] \end{aligned} \tag{8.15}$$

where $m(x) = E\{Z(x)\}$ is the known expected value, not necessarily stationary, of the random variable $Z(x)$. The $M+j-1$ weights a_k in 8.15 are obtained from the covariance function C, which is not necessarily stationary, solving the simple kriging normal system:

$$\sum_{k=1}^{M+j-1} a_k C(x_k, x_i) = C(x_j, x_i), \qquad i = 1, M + j - 1 \tag{8.16}$$

The (conditional) variance of the random variable $Z(x_j)$ is identified with the simple kriging variance:

$$VAR\{Z(x_j) \mid z(x_k), k = 1,.., M + j - 1\} = C(x_j, x_j) - \sum_{k=1}^{M+j-1} a_k C(x_j, x_k) \tag{8.17}$$

The SGSIM is extremely convenient because it only requires the estimation of two parameters to determine the distribution at each point. On the other hand, the assumption of a multi-variate Gaussian random field required by this method is difficult to test, and it is often not justified in reality (Caers, 2000). One approach would be to transform the sampling distribution into a Gaussian distribution through the use of the normal score transformation, which for a random variable Z with cdf $F_Z(z)$ is defined as:

$$Y = G^{-1}(F_Z(Z)) \tag{8.18}$$

where G^{-1} is the inverse cumulative density function of the standard normal distribution. If Z is a sample distribution, the normal-score transformation is accomplished by sorting the elements of the distribution in ascending order, creating the sample cdf, and then applying equation 8.18 to obtain the normal score values. The normal score transformation defined in this way is then applied directly or inversely by interpolating the tabulated values.

A drawback of this approach is that the normal-score transformation may alter the relation between the primary variable and the secondary variables that may be used to deriver the distribution of the primary mean. In addition, SGSIM will reproduce the normal-score variogram, which if the original distribution is markedly skewed, does not automatically translate in a good estimate of the original variogram (Caers, 2000).

8.5.4.3 Sequential Indicator Simulation

When the random field to be simulated cannot be successfully transformed into a multivariate normal distribution, SISIM may provide a framework for simulating the random field in a non-parametric way. In SISIM, step 3.a is executed substituting the estimate of the mean and variance of the conditional normal distribution (equations 8.15 – 8.17) with a non-parametric indicator kriging approach. Indicators are random variables defined as:

$i(k;x)$ = 1 if $z(x)$ belongs to category k, 0 otherwise for categorical variables
$i(z_k;x)$ = 1 if $z(x)<=z_k$, 0 otherwise for continuous variables (8.19)

and having the following useful property:

$E\{i(k;x)\}$ = Prob$\{Z(x)$ ε category $K\}$ for categorical variables
$E\{i(z_k;x)\}$ = Prob$\{Z(x)<=z_k\} = F_z(z_k)\}$ for continuous variables (8.20)

This last property extends also to the conditional case, providing the basis for non-parametric least-square estimation of the conditional cdf of $Z(x)$. In the continuous case, the indicator kriging approach subdivides the range of the random field Z in $K+1$ intervals, delimited by the cutoff thresholds z_k, $k=1, \ldots, K$. For each point x of the region of interest and threshold z_k, the indicator $i(z_k;x)$ is thus a binomial random variable, with

expected value equal to the conditional probability that $P(Z(x))$ is less than or equal to z_k. If the random field Z is stationary, then $E\{i(z_k;x)\} = F_z(x,z_k) = F_z(z_k)$ and equations 8.12 - 8.14 can be transformed for estimating the indicator $i(z_k;x)$, given $z(x1),\ldots ,z(xM)$ and the knowledge of $F_z(z_k)$:

$$i^*(z_k;x) = F_Z(z_k) + \sum_{j=1}^{M} a_j(z_k,x)(i(z_k;x_j) - F_Z(z_k))$$

$$\sum_{l=1}^{M} a_l(z_k,x)C_I(z_k;x_l - x_j) = C_I(z_k;x - x_j) \qquad \text{for } j = 1,\ldots,n \tag{8.21}$$

where: $i^*(z_k;x)$ = estimator of $i(z_k;x)$

$C_I(z_k;x - y)$ = Stationary covariance between $i(z_k;x)$ and $i(z_k;y)$

Note that $i^*(z_k;x)$ takes any value between zero and one and its value is taken as a estimate of $F(z_k|z(x_1),\ldots,z(x_M))$. For each point x, then, the conditional cdf is reconstructed by adequately interpolating the K estimates $F(z_k|z(x_1),\ldots,z(x_M))$ obtained from applying equation 8.21 to the K thresholds z_k, $k=1, \ldots K$. This conditional cdf is then used in step 3.b of the sequential simulation algorithm to produce one realization of the random field at x. The number K of thresholds used to reconstruct the random field cdf must be carefully selected to balance the desired precision in reconstructing the cdf, computational cost, nature of the random field to be simulated, and the variability of the indicator covariance function. A common approach is to select as thresholds the separators between the deciles of the global cdf distribution (Isaaks and Srivastava, 1989; Atkinson and Lloyd, 1998).

SISIM is a very flexible technique that allows estimating distributions that cannot be transformed into Gaussian distributions, even mixed distributions (Glacken and Blackney, 1998). Notwithstanding, SISIM has several limitations:

- The most obvious SISIM drawback is that it requires solving the normal system of equations 8.20 K times for each location *x*, instead of just once, as is the case of SGSIM, increasing the computation almost by a factor of *K*.
- The *K* indicator variables are computed independently from each other, producing values that are often inconsistent with the requirements of a cdf (i.e., $i(z_k,x) <= i(z_{k+1},x)$, $0<= i(z_k,x)<=1$). The most common approach to this problem is to correct the sequence of indicators produced by the kriging to ensure consistency with the cdf requirements (Deutsch and Journel, 1998). However, this action decreases the capability to reproduce the desired correlation and may add some bias.
- In many instances, the cumulative probability distributions in the lowermost and uppermost classes are not linear, requiring special handling (Glacken and Blackney, 1998).
- As mentioned above, the choice of the thresholds separating the various probability classes is not always straightforward.
- Despite the more cumbersome approach, indicator kriging does not guarantee a better representation of data variability than ordinary kriging (Atkinsons and Lloyd, 1998).
- Up to *K* different covariance functions may have to be formulated.
- SISIM will reproduce the variogram of each indicator, which does not automatically translate in a good estimator of the original variogram (Caers, 2000).

8.5.4.4 Direct Sequential Simulation

As mentioned before, equations 8.11 – 8.14 estimate the mean and variance of a random variable given some measured values. However, with the exception of a few

cases (e.g., Gaussian distributions) mean and variance are not sufficient for complete distribution characterization (Soares, 2001). On the other hand, the transformations required for using the SGSIM or SISIM procedures – if possible – may prevent the estimation of the original covariance. DSSIM is a set of techniques, still in the developmental phase, aiming to simulate a random field without transforming the original variables and back-transforming the simulation results. Caers (2000) proposed to reconstruct the conditional cdf at *K* predefined thresholds by imposing them to satisfy a series of constraints:

1. Cdf consistency constraints

$F(z_k,x|M) <= F(z_{k+1},x|M)$, $0<= F(z_k,x|M)<=1$

2. Compatibility with the estimated simple kriging mean and variance

$$Z^*_{SK}(x) = \sum_{k=1}^{K} \frac{t_{k+1}+t_k}{2}(F(z_{K+1},x\,|\,M) - F(z_K,x\,|\,M)$$

$$\sigma^2_{SK}(x) + Z^*_{SK}(x)^2 = \sum_{k=1}^{K}\left(\frac{t_{k+1}+t_k}{2}\right)^2 (F(z_{K+1},x\,|\,M) - F(z_K,x\,|\,M)$$

3. Consistency with the global cdf $F(z_k)$. This is pursued by minimizing the objective function:

$$O = \sum_{k=1}^{K}\left(\frac{F(z_K,x\,|\,M) - F(z_K)}{F(z_K)}\right)^2$$

Soares (2001) suggested to generate the conditional cdf drawing data from the intervals of the global distribution with center around the simple kriging mean and range depending on the simple kriging variance. Oz et al (2003) used the inverse of the global normal score transformation to find the shape of the local conditional cdf corresponding

to the Normal distribution with the estimated conditional mean and variance. The cdf determined in this way is then sampled to produce the desired realizations. Neither the conditional data nor the simulation results are transformed into Gaussian variables during the process.

8.5.5 Selection of the Correlated Random Field Generator for Remote Sensing of Precipitation

The challenge in integrating the remote sensing information in the generation of a spatially correlated random field is that it changes the local unconditional precipitation distribution on a pixel-by-pixel basis. This means that, at the very least, the random field to be simulated does not have a stationary mean and possibly not even a stationary covariance.

The distribution of the TRMM rain rates used to estimate precipitation in the empirical distribution approach is too skewed and presents too many no-rain samples to be considered normal even after the normal score transformation (Figure 8.13). Therefore, the generation of a spatially correlated precipitation field based on the empirical distributions can be achieved only by the SISIM or DSSIM approaches. DSSIM, however, is still in a developmental phase and lacks an established computational procedure, narrowing the set of possible choices to the SISIM procedure alone. Since this non-parametric approach can accommodate spatially varying random fields, the only potential drawback is its higher computational burden.

The approach by Bell (1987) is a natural match for the analytical-distribution model. Unfortunately, that model requires that only one probability of rain/no-rain and one precipitation intensity distribution is used for reconstructing the covariance to be imposed to the Gaussian distribution from the covariance of the precipitation model.

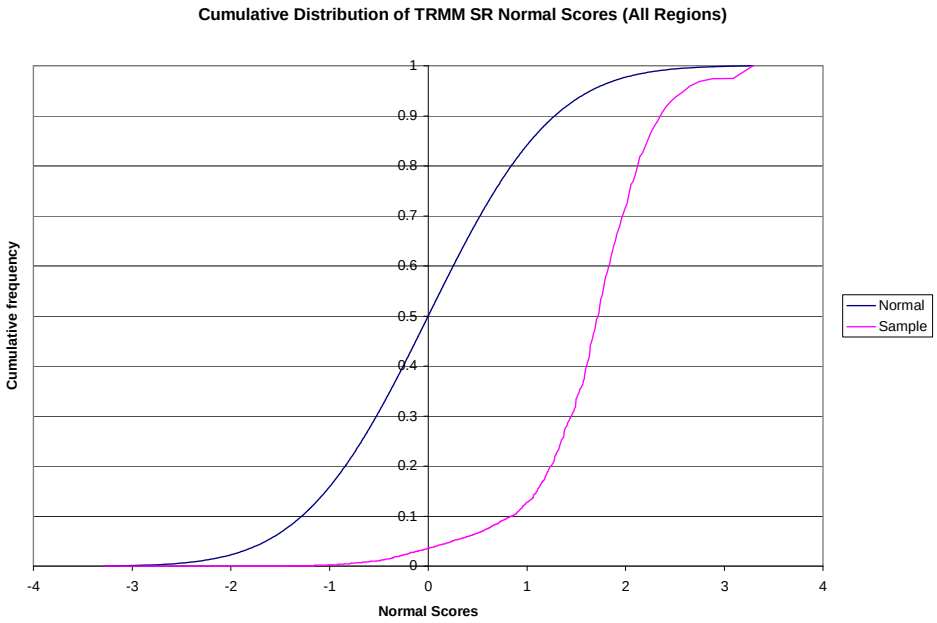


Figure 8.13 Normal score distribution of the 1998-1999 TRMM SR samples.

Doing this, however, would imply that a significant part of the information provided by the satellites is ignored. Furthermore, the approach may not be computationally very efficient since it requires the simulation of the entire domain, even when the precipitation is restricted to small sectors.

The logical alternative is to simulate the rain/no-rain areas first and to reserve the simulation of the rain intensity only for the rainy pixels. This approach is generically recommended by Deutsch and Journel (1998), who suggest a two-step approach when the random field consists of a mixture of different physical and/or statistical populations. In the first step, the geometry of the mixture should be simulated, while the simulation of the population attributes should be addressed later, when it is possible to perform it within homogeneous populations. A similar approach was also taken by Barancourt et al. (1992) who used it to explicitly consider precipitation variability due to fractional coverage of rainfall in the interpolation of rain gage data. In their work, the first step defines the rainy area using a one-class indicator kriging and conditioned by the rain gage

showing rain or no-rain. In the second step, the difference between the precipitation intensity and the event average precipitation is modeled as Gaussian. The standard kriging is then performed for every pixel, conditioned using only rainy gages. The key assumption in this method is that the rain intensity of rainy pixels is independent from the intermittency. The authors support this assumption showing that the structure and mean of the precipitation is not related to the size of the rainy area and that the precipitation shows little change with distance from the border of the rainy area. Pardo-Iguzquiza et al. (2006) states that this assumption is reasonable at least for the cases where the scale of precipitation variability is much smaller than the spatial extent of the rainy area. A shortcoming of this procedure is that it does not directly estimate the variance of the overall precipitation rate and consequently cannot be used in an estimation context, but requires the simulation of an ensemble of values. Seo (1998a, 1998b) developed some relations involving the conditional expectation of precipitation occurrence and amount, to provide the comprehensive estimation variance. Following the same approach, Pardo-Iguzquiza et al. (2006) also show a method for determining the estimate variability, but also caution that the non-Gaussian nature of the overall distribution reduces its applicability. They argue that it is preferable to use the simulation for building confidence intervals for the rain estimates.

Theoretically, the simulation of the lognormal distribution of the rain intensities could be done also with spectral methods or matrix decomposition. However, since the sequential simulation approach will be taken for estimating the rain/no-rain distribution, it is more logical to use it also for estimating the precipitation rate. Furthermore, the SGSIM paradigm is more flexible in dealing with spatially varying distributions of rain rates.

Therefore, the spatial correlation of the analytical-distribution model is accomplished as a two-step approach, using a one-threshold SISIM procedure for

discriminating rain/no-rain pixels and a SGSIM procedure for simulating the precipitation intensity.

8.6 A Sequential Simulation Approach to Remote Sensing of Precipitation

The development of the SISIM and mixed SISIM/SGSIM procedures that are used herein to generate the spatially correlated precipitation fields using respectively the empirical and analytical distributions follows the steps noted below:

1. Development of covariance function models in the two procedures.
2. Simulation of single instantaneous precipitation distributions for a set of representative cases and comparison with the corresponding TRMM images.
3. Simulation of temporally uncorrelated precipitation.
4. Simulation of temporally correlated precipitation. In this case, the temporally correlated satellite distributions are used as priori estimates for the sequential simulation algorithms.

The software implementing the sequential simulation procedures integrates the appropriate subroutines provided by the Geostatistical Software LIBrary (Deutsch and Journel, 1998) into the programs written for the single pixel simulation. Although the GSLIB uses a regular rectangular grid for the simulation, the sequential simulation methods do not need to be applied to all the pixels of the grid. This allows restricting the sequential simulation to the pixels that are potentially rainy (namely, the pixels having a prior probability of rain greater than zero), greatly decreasing the computation time. In Addition, the pixels for which the prior probability of rain is zero are not used for conditioning the sequential simulation, since they are no longer considered random. Thus, the covariance functions used by SISIM and SGSIM have been computed using only potentially rainy pixels.

Another aspect to consider is that, at least in this version of the method implementation, no precipitation data is used as "measured" values for constraining the precipitation field at any given instant. This means that from the simulation is conditioned only by "soft" data and not by "hard" data.

Although the EKJ area where the multi-pixel analysis is performed is 1°x1°, corresponding to a grid of 20x20 pixels, the spatially correlated simulation has been performed over a 50x50 grid concentric with the EKJ area. The larger simulation area allows accounting for the large majority of the external influence on the EKJ pixels, some effects may remain unaccounted.

8.6.1 Empirical Distributions, Sequential Indicator Simulation

As described in section 8.5.4.3, SISIM partitions the range of the random field to be simulated in a series of intervals separated by cutoff thresholds. For each threshold, SISIM simulates the corresponding indicator random field as defined in equations 8.21. In the present case, however, the unconditional precipitation cdf $F_R(z)$ is not fixed in space, but varies according to the satellite information. This means that equation 8.21 must be changed as follows:

$$i^*(z_k;x) = F_R(z_k;x) + \sum_{j=1}^{M} a_j(z_k,x)(i(z_k;x_j) - F_R(z_k;x)) \qquad (8.22)$$

Deutsch and Journel (1998) state that it is possible to use the simple kriging algorithm for computing 8.22 as long as (1) $F_R(z_k;x)$ is known for every point x of the grid and every cutoff z_k, and (2) the indicator residual covariance (that is the correlation of the residuals $i(z_k;x) - F_R(z_k;x)$) is used in place of the indicator covariance. They also recommend that the correlation be stationary and the prior $F_R(z_k;x)$ be smooth.

While it is possible to assume that the indicator covariance is stationary, the prior $F_R(z_k;x)$ at the 5 km resolution does not always vary smoothly, even when one considers

only potentially raining pixels. For example, strong spatial variations in $F_R(z_k;x)$ are very likely at the leading edge of the storm.

Equation 8.22 states that the conditional cdf at pixel x and cutoff z_k (for example 30 mm h^{-1}) is given by the unconditional cdf plus a linear combination of the differences between the actual probability that the precipitation is below z_k and the respective unconditional cdf at some neighboring point. This relation, however, does not consider the pixel precipitation type. For example, let x be a pixel under stratiform conditions and y a neighboring pixel under convective conditions. The probability that precipitation intensity at x is above 30 mm h^{-1} is zero (that is $F_R(30;x)$ =1.0), while the probability of having rain of such intensity at y is substantial (e.g., $F_R(30;y)$ = 0.8). However, if y is the conditioning element and the actual precipitation is larger than 30 mm h^{-1}, then the conditional cdf at x will be less than 1.0, allowing the possibility of such intense rain, even if this event is not possible. Reversely, if x is the conditioning element, the fact that its precipitation is never above 30 mm h^{-1} will have an undue impact on the rain distribution of y.

To avoid this problem, relative rather than absolute thresholds are used herein. This means that instead of using a series of thresholds such as 0, 2, 5, 10 mm h^{-1}, etc., the indicator kriging is computed for thresholds 0, 0.1, 0.2, etc.. These relative thresholds are obtained as the ratios between the simulated rain and the maximum value of the local unconditional distribution. In other words, the precipitation intensity is rescaled according to the observed maximum precipitation for the pixel conditions.

With this change, Equation 8.22 gives the probability that at pixel x the precipitation is below, say, 30% of its range as the sum of the unconditional cdf for that level and a linear combination of the differences between the actual probability that the precipitation is below 30% of its range and the corresponding unconditional cdf at some neighboring point.

Essentially, this formulation implies that the chance the precipitation belongs to a certain quantile of the local precipitation range is conditioned by the precipitation in neighboring pixels falling in the same quantile of their own local distributions. This approach also has the advantage of smoothing the distribution of $F_R(Z;x)$ and increasing the stationarity of the indicator residual covariance.

After extensive experimentation, the thresholds chosen for the precipitation simulation were set to 0.0, 0.1, 0.2, 0.3, 0.4, 0.6, and 0.8, while the cdf at 1.0 is set to 1.0. The higher resolution at the lower end of the precipitation range is chosen to better model the skewed distribution of precipitation.

8.6.1.1 Modeling of the Indicator Residual Covariance

As mentioned in Section 8.5.4.1, the normal system of equations 8.12 has a unique solution only if the covariance matrix K is positive definite. This means that the covariance cannot be computed just by interpolating some sample values, but must be produced using specific functional types. The approach adopted in GSLIB is to model the variogram combining a small number of basic analytical functions that certainly provide a positive definite matrix and to use this analytical version of the variogram to compute the covariance structure of the random field to be simulated. The variogram is defined as half the average squared difference between two attribute values separated by vector **h** (Deutsch and Journel, 1998):

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(h)} (z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h}))^2 \qquad (8.23)$$

Where: **h** = vector separating the samples

N(**h**) = number of elements separated by vector **h**.

In addition, the following relations link the variogram with the covariance and correlation coefficient:

$$C_z(\mathbf{h}) = \gamma(\infty) - \gamma(\mathbf{h}), \qquad C_z(0) = \gamma(\infty), \qquad \rho(\mathbf{h}, z) = \frac{\gamma(\infty) - \gamma(\mathbf{h})}{\gamma(\infty)} \tag{8.24}$$

The indicator residual covariance may assume different shapes depending on the value of the threshold z_k. Thresholds 0.0 and 0.1 feature both a short-range and a long-range component, while the higher thresholds could be well represented with just one component. Deutsch and Journel (1998), however, report that using the same combination of basic models for all the thresholds is a way to decrease the number of cdf inconsistencies during the simulation. Thus, the following model is used for all thresholds, where a_1 accounts for the shorter-range component and a_2 for the longer-range component:

$$\gamma(h) = n + c_1 \cdot \left[1 - \exp\left(-\frac{3h}{a_1}\right)\right] + c_2 \cdot \left[1 - \exp\left(-\frac{3h}{a_2}\right)\right] \tag{8.25}$$

As depicted in Figure 8.4, extensive testing has shown that the average indicator residual covariance can be approximated as isotropic, allowing the use of the same set of parameters for all directions (Table 8.1).

Table 8.1, Figure 8.14, and Figure 8.15 reveal that the correlation between indicator residuals is relatively strong for $z_k = 0.0$ (rain/no-rain), $z_k = 0.1$, and z_k=0.2. At higher thresholds, the correlation decreases rapidly, in strength as well as range, which is understandable given the skewed shape of the precipitation distribution. In this last case, the range of the long-distance component varies significantly and without a clear pattern, mainly due to their small significance (ratio $a_1/a_2 > 10$). Furthermore, the indicator

residual correlation for lake and land pixels are much more similar to each other than the correlation of the precipitation alone (Figure 8.2).

Table 8.1 The coefficients of the variogram models used in the SISIM simulation.

Threshold	0.0	0.1	0.2	0.3	0.4	0.6	0.8
Lake							
n	0.0014	0.0084	0.0088	0.0076	0.0074	0.0053	0
c1	0.053	0.0575	0.0438	0.0328	0.0228	0.0105	0.0092
a1 (km)	19.7	26.2	26.6	23.50	21.3	13.4	7.6
c2	0.0712	0.0240	0.0075	0.0034	0.0016	0.0010	0.0004
a2 (km)	108.4	126.9	111.8	321.9	327.3	117.6	238.5
Land							
n	0.005	0.005	0.003	0.0012	0	0	0
c1	0.0587	0.0430	0.0306	0.0233	0.0183	0.0109	0.0067
a1 (km)	25.1	20.5	16.7	14.9	12.7	9.4	7.7
c2	0.038	0.0104	0.0039	0.0016	0.001	0.0004	0.0002
a2 (km)	136.6	203.4	219.7	247.5	220.0	165.0	110.0

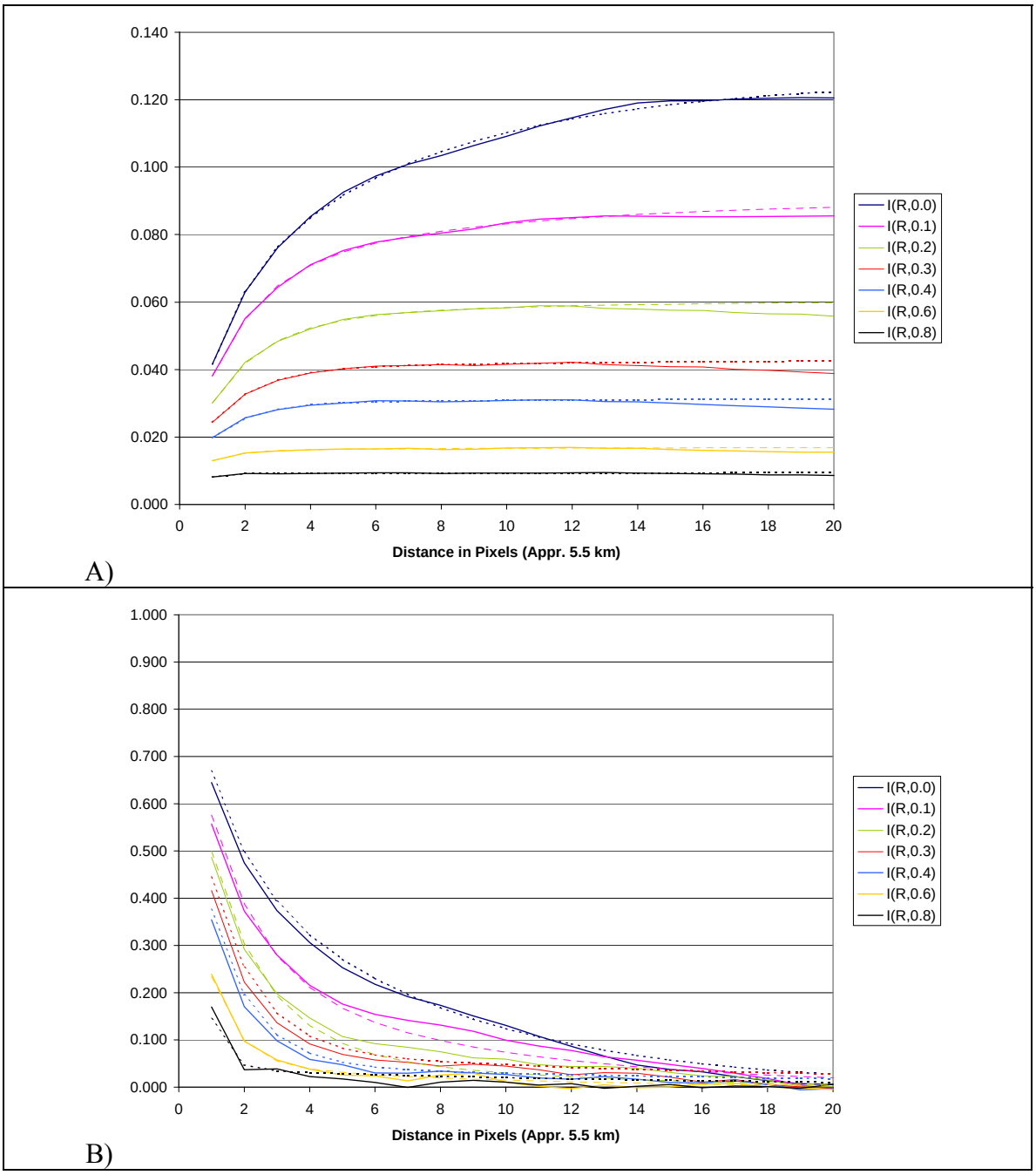


Figure 8.14 Sample (solid) and model (dashed) indicator residual variogram (A) and correlation (B) for lake pixels.

Figure 8.14 and Figure 8.15 show that the model represents well both the sample variograms and the sample correlation, although in several instances the model is a few percentage points below the sample correlation. In addition, the figure illustrates that

indicator residuals of instantaneous precipitation have very little correlation for distances above 55-60 km, similar to the behavior of instantaneous precipitation itself.

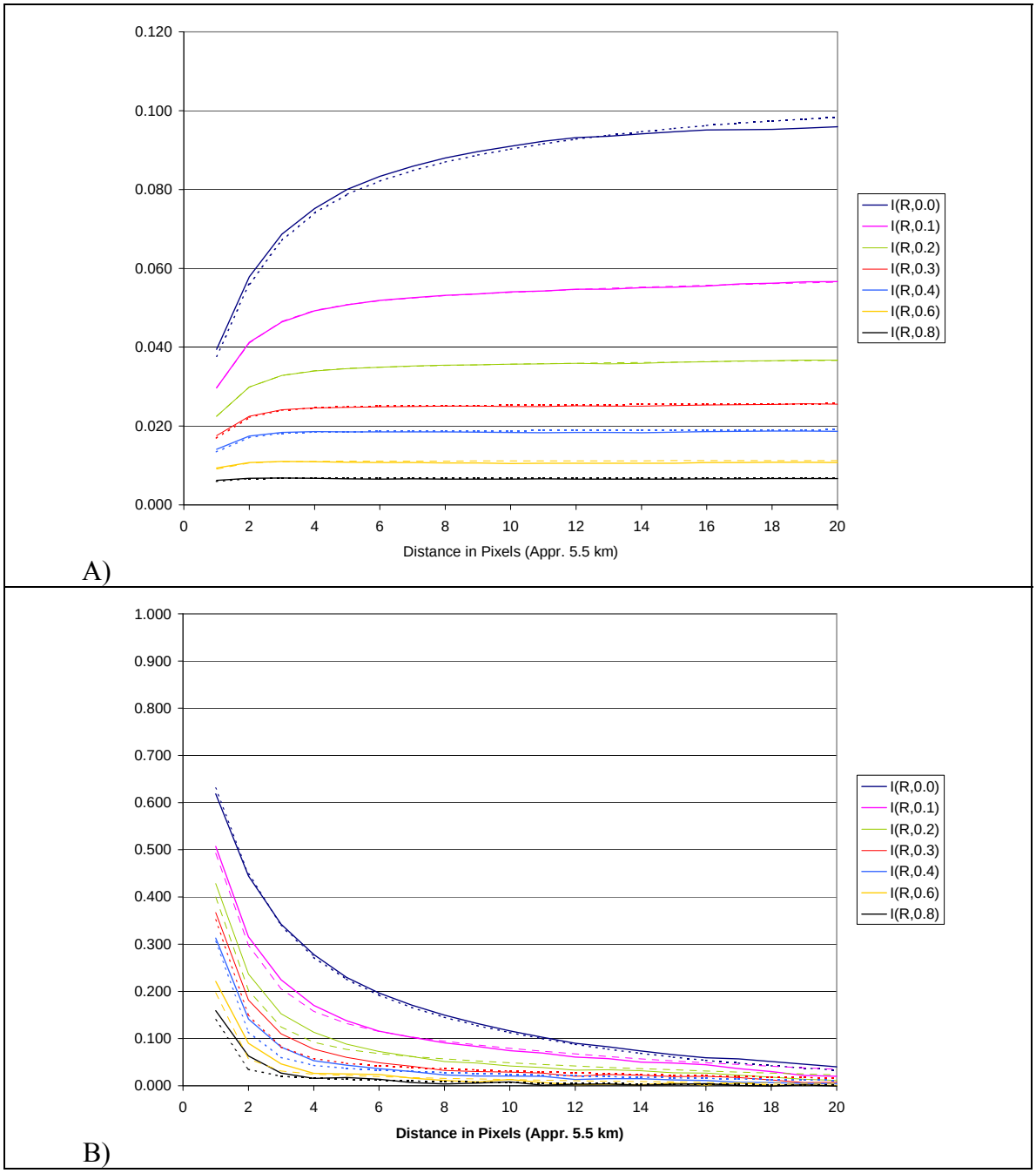


Figure 8.15 Sample (solid) and model (dashed) indicator residual variogram (A) and correlation (B) for land pixels.

8.6.1.2 Comparison with single TRMM Passes

The sequential simulation technique is used to create correlated random fields of half-hour precipitation values that are accumulated over a certain period to produce daily, dekadal, monthly, etc., precipitation estimates. Consequently, the behavior of the procedure at this basic level is being examined first by comparing temporally uncorrelated precipitation against single TRMM images representative of small and large storms over the entire basin.

The simulation presented in Figure 8.16 - Figure 8.19 was obtained by applying the SISIM technique to each potentially rainy pixel using the satellite derived cdf as prior, means with relative thresholds, and covariance functions derived in section 8.6.1.1. The maximum number of pixels used in the kriging estimation was twelve.

The SISIM realizations in Figure 8.16 B) to Figure 8.19 B) show much higher continuity in space than those produced by the spatially uncorrelated precipitation (Figure 8.6). On the other hand, single spatially correlated realizations may completely ignore some rainy areas, while the uncorrelated realizations always include at least some rainy pixels in large rainy areas.

In Figure 8.16 C) to Figure 8.19 C) the sample precipitation distribution (solid lines) is compared against the precipitation distribution for the entire ensemble of realizations (dashed lines) as a function of the storm stage (the label -3 is assigned to unclassifiable pixels and the label -2 is assigned to the pixels not belonging to a convective storm). It can be seen that the model reproduces relatively well the observed distributions, although it may somewhat overestimate the average precipitation of the uncorrelated precipitation (Table 8.2).

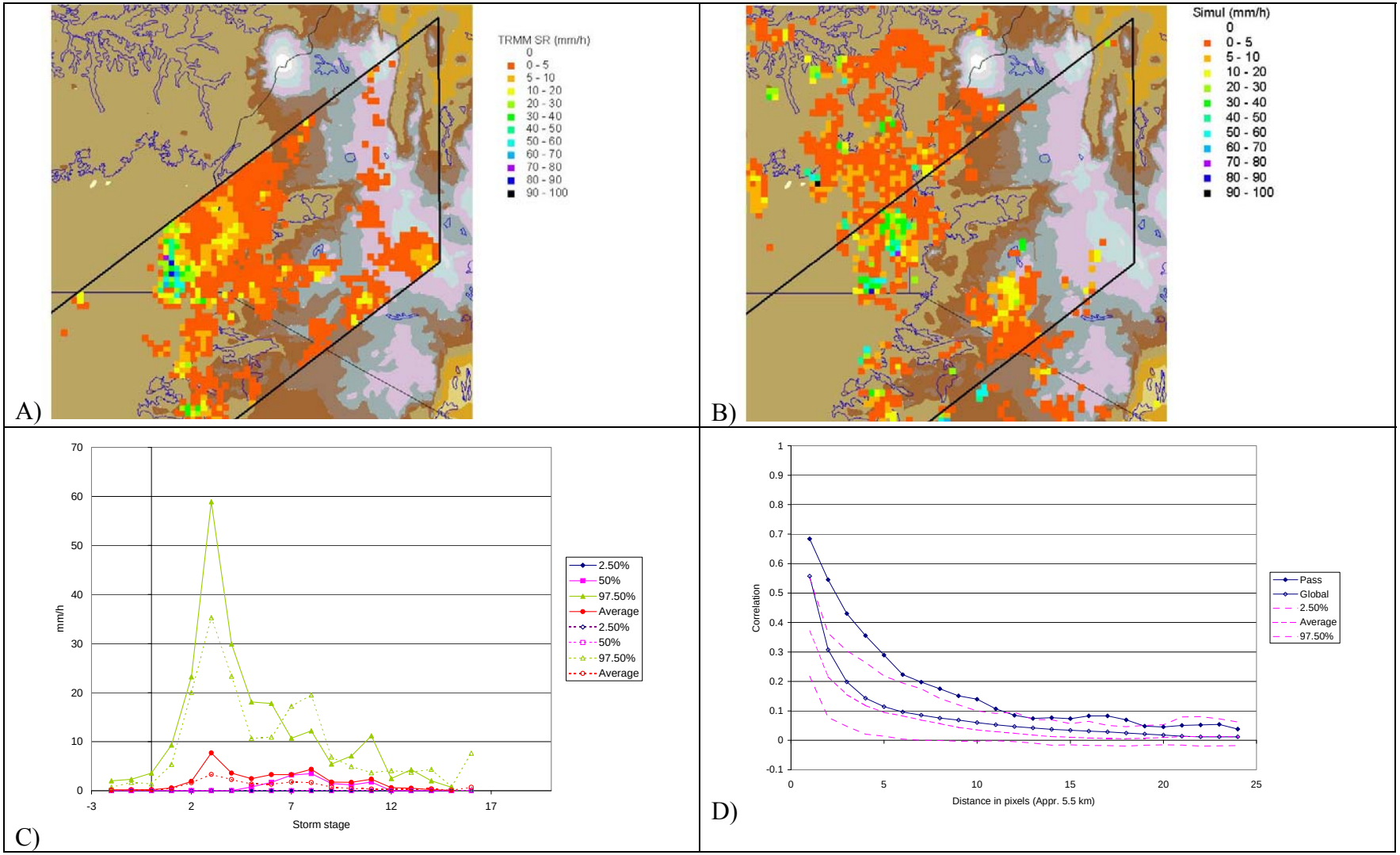


Figure 8.16 Simulation of single TRMM passes. A) Observed TRMM SR map; B) One realization of the TRMM rain map; C) Observed and model precipitation distribution. D) Observed and model spatial correlation. Jan 09, 1998 22:12 GMT

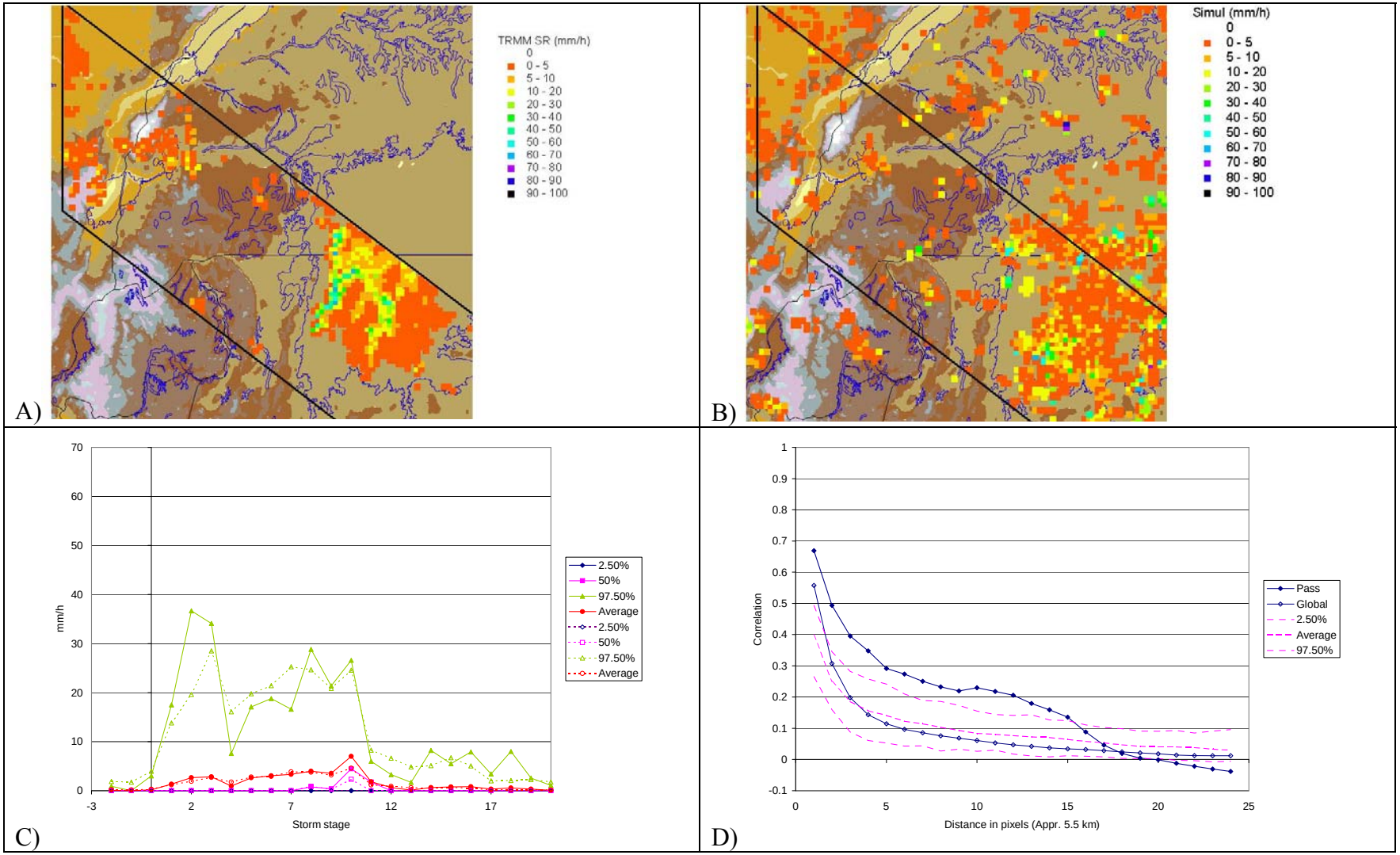


Figure 8.17 Simulation of single TRMM passes. A) Observed TRMM SR map; B) One realization of the TRMM rain map; C) Observed and model precipitation distribution. D) Observed and model spatial correlation. May 3, 1998 23:26 GMT.

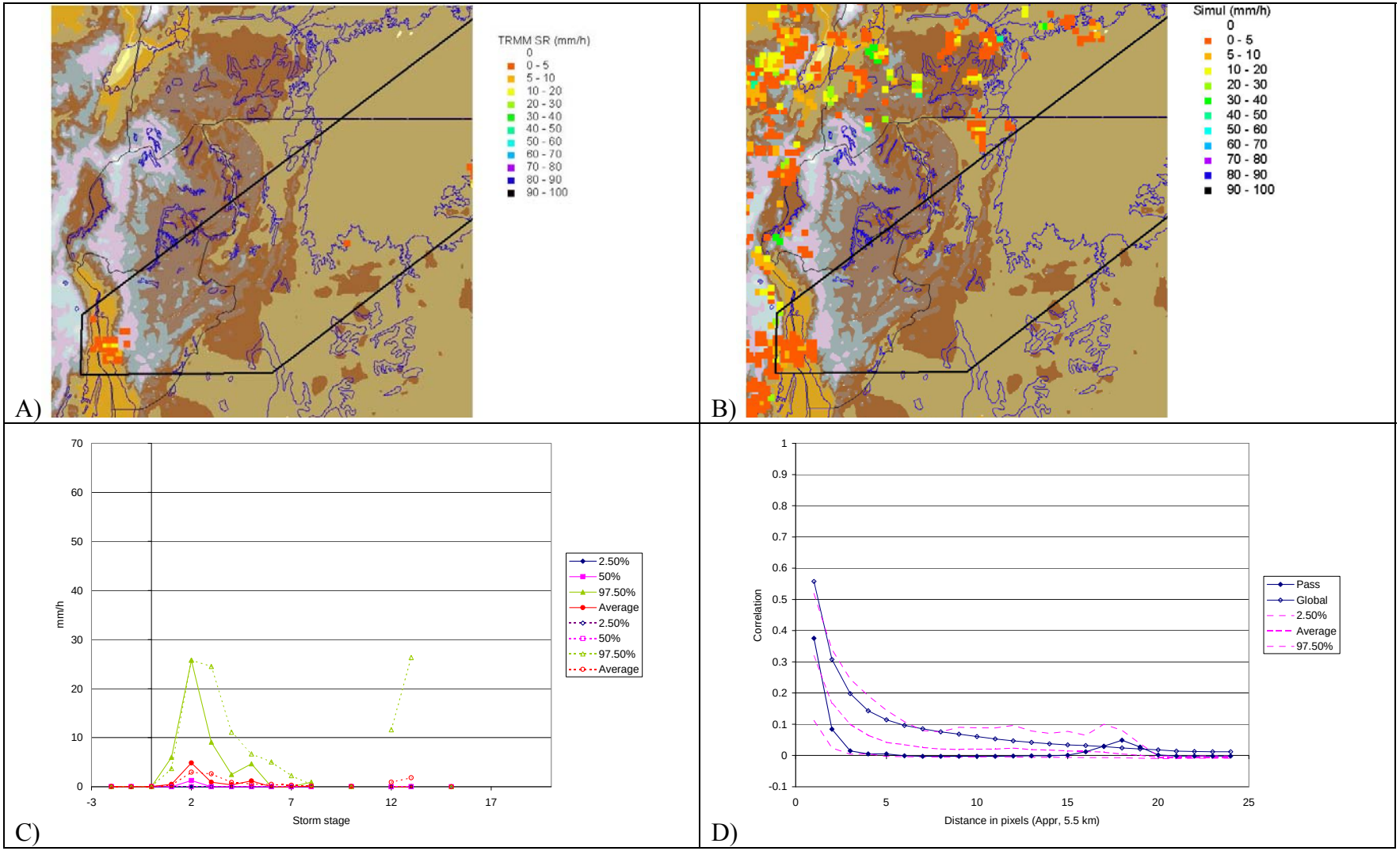


Figure 8.18 Simulation of single TRMM passes. A) Observed TRMM SR map; B) One realization of the TRMM rain map; C) Observed and model precipitation distribution. D) Observed and model spatial correlation. May 04, 1998 10:50 GMT.

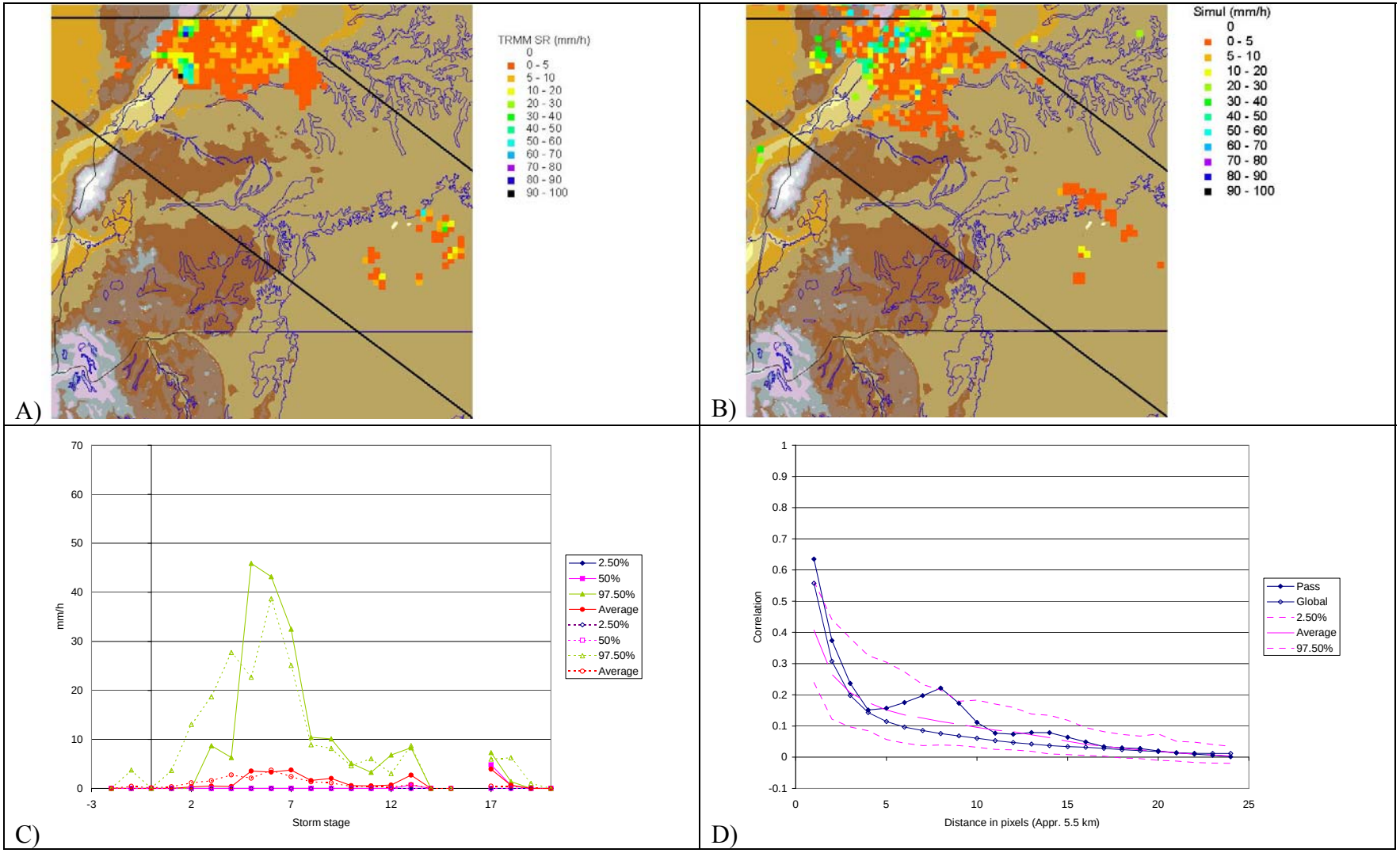


Figure 8.19 Simulation of single TRMM passes. A) Observed TRMM SR map; B) One realization of the TRMM rain map; C) Observed and model precipitation distribution. D) Observed and model spatial correlation. July 17, 1998 08:40 GMT.

The set of figures from Figure 8.16 D) - Figure 8.19 D) compares the sample spatial correlation (solid lines) with the average spatial correlation of the single-realization spatial correlations (dashed lines) as well as the correlation range.. It is evident that the model replicates the spatial correlation of small storms, but not the spatial correlation of larger, more organized storms.

Furthermore, in these comparisons, it is not able to reach the average short-distance correlation. To understand whether the indicator residual correlation is not representing the actual residual correlation or whether this result is due to other reasons, the TRMM relative indicator residual spatial correlation for the levels 0.0, 0.1, 0.2, and 0.3 is reported in Figure 8.20 for the May 3, 1998 23:26 GMT TRMM pass. The charts also report the residuals autocorrelation functions for land and lake used in computing the kriging estimates, as well as the actual residual correlation produced by kriging.

The match between the observed TRMM correlation and the two desired correlation functions is good, with the lake function performing slightly better probably because this TRMM swath is dominated by a large storm over Lake Victoria. On the other hand, the spatial autocorrelation resulting from the kriging matches the desired and sample correlation well for the indicators above 0.2, approximately for the 0.1 indicator, and poorly for the 0.0 indicator. It is noted that indicators 0.0 and 0.1 have the highest covariance, which in areas of enhanced variability results in large violations of the prior cdf consistency constraints. The corrections necessary to meet the consistency constraints cause substantial deviations from the theoretical relation of equation 8.22, particularly at short distances where the highest covariance creates the largest deviations.

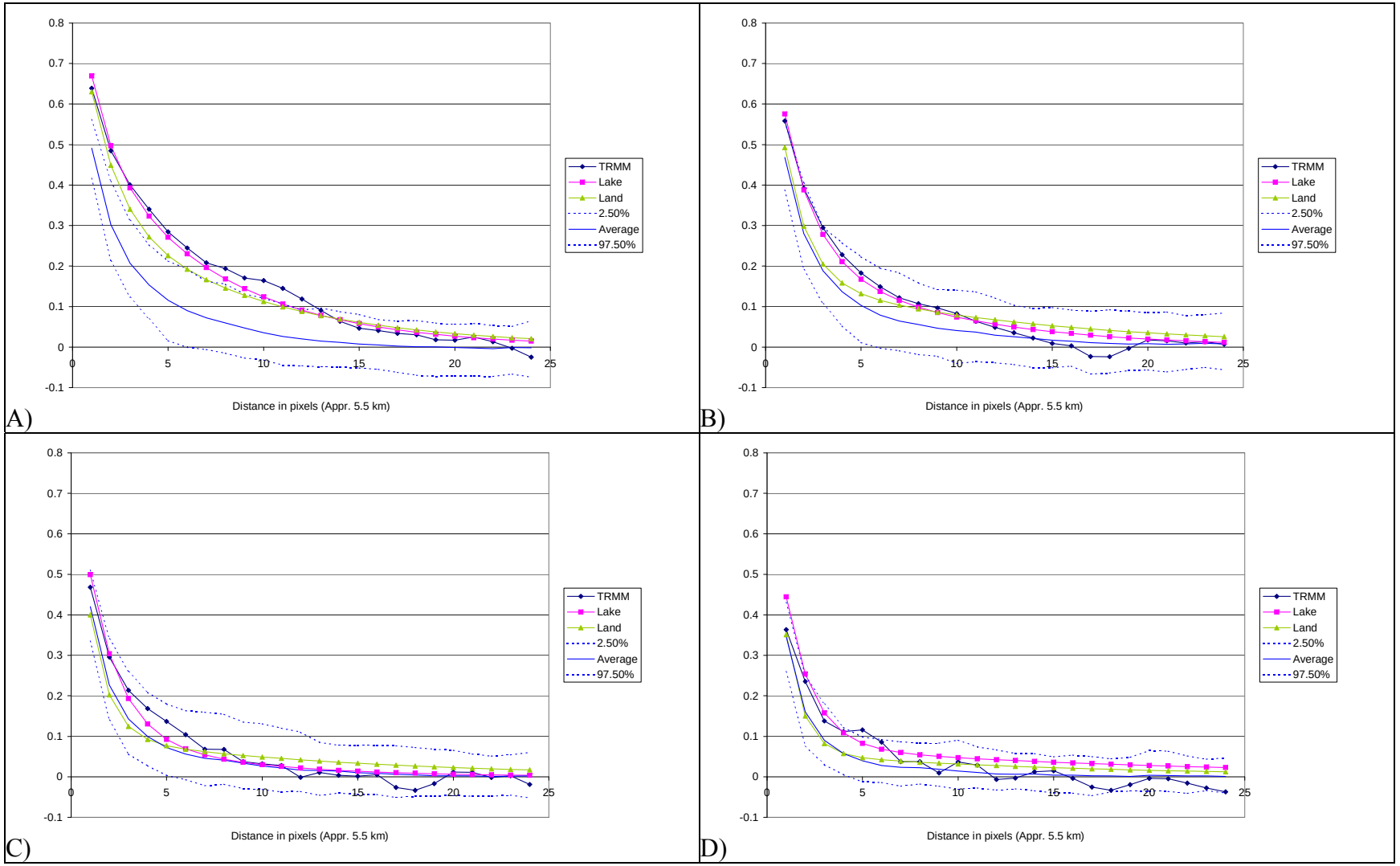


Figure 8.20 TRMM, model, and actual relative indicator residual spatial correlation. May 3, 1998 23:26 GMT. A) 0.0 Level; B) 0.1 Level; c) 0.2 Level; D) 0.4 Level

The best way to address this problem would be to decrease the spatial resolution of the simulation from the single pixel (5.5 km) to 2x2 (11 km) or 3x3 (16 km). This would allow a smoother distribution of the prior cdf and less skewed distributions, decreasing the number of cdf violations. It would also increase the residual indicator spatial correlation and probably reduce the noise caused by the precipitation temporal correlation noted in section 8.4. Unfortunately, it was not possible to perform this change in this study because 2x2 or 3x3 MAP cannot be compared against single gages, especially at the daily and ten days levels, due to the strong difference in variability caused by averaging.

Furthermore, the density of rain gages is not sufficient to estimate 2x2 or 3x3 MAP anywhere in the basin. An additional complication is that it would not be possible to correctly estimate the 3x3 precipitation temporal correlation using the technique presented in Section 7.4, because 3x3 areas centered in neighboring pixels at consecutive stages would have more than 40% overlap.

The procedure adopted in this study is to maintain single pixel spatial resolution, but at the same time simulate a 3x3 behavior by the following two-step procedure:

1. Kriging is performed as in equation 8.22 for estimating the conditional cdf at a point x based on the neighboring indicator residuals.
2. The cdf modifications produced by kriging and the sampling probability used for determining the realization of the random field in x are extended to the eight pixels adjacent to x, making the nine pixels behaving as a 3x3 area.

Figure 8.21 shows that the proposed procedure, named F-extension, substantially improves the residual indicator dynamics at the lower levels, without affecting it too much at the higher levels. At medium range, however, the residual indicator dynamics is still below the desired level.

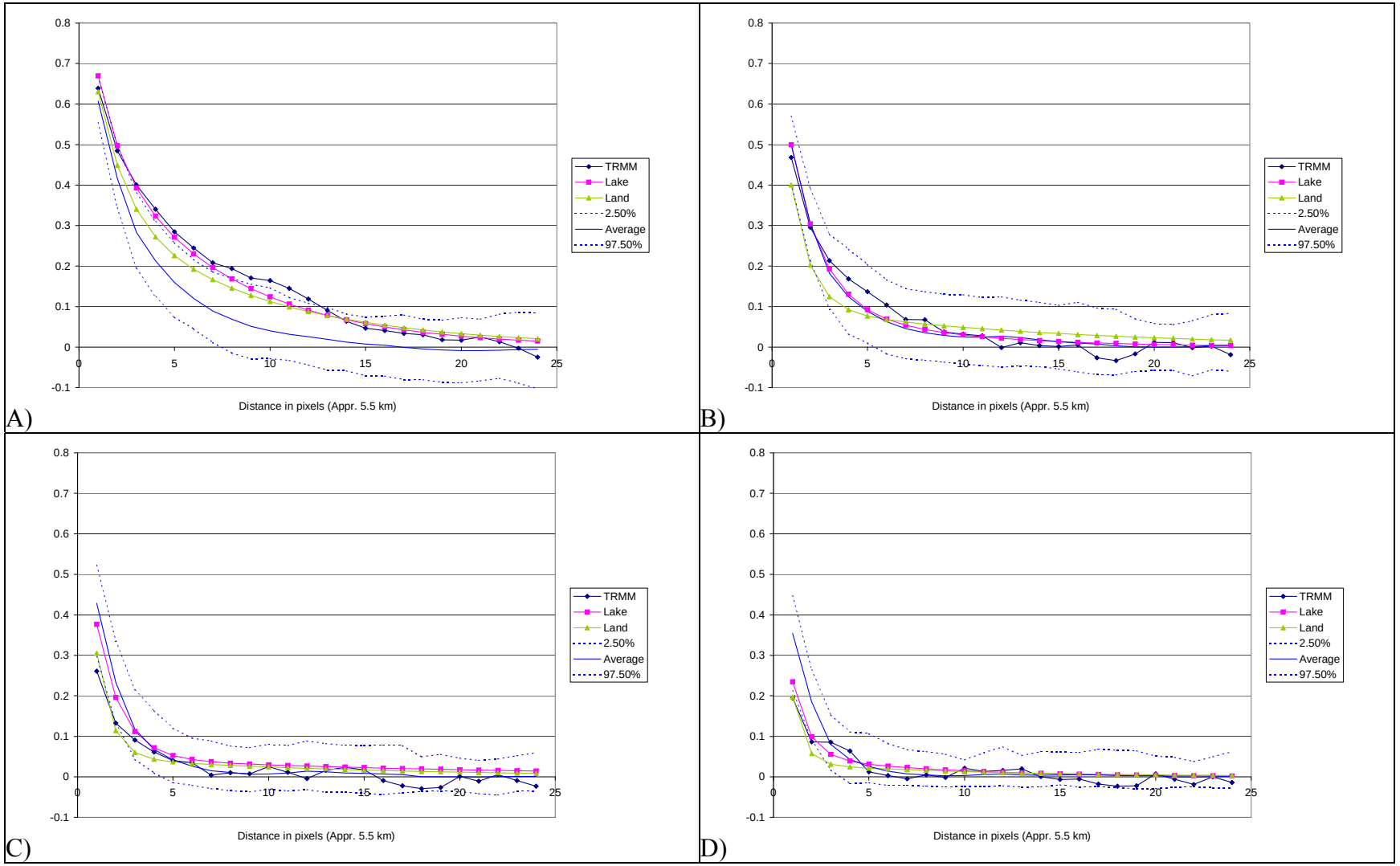


Figure 8.21 TRMM, model, and actual relative indicator residual spatial correlation for F-extension approach. May 3, 1998 23:26 GMT. A) 0.0 Level; B) 0.2 Level; c) 0.4 Level; D) 0.6 Level

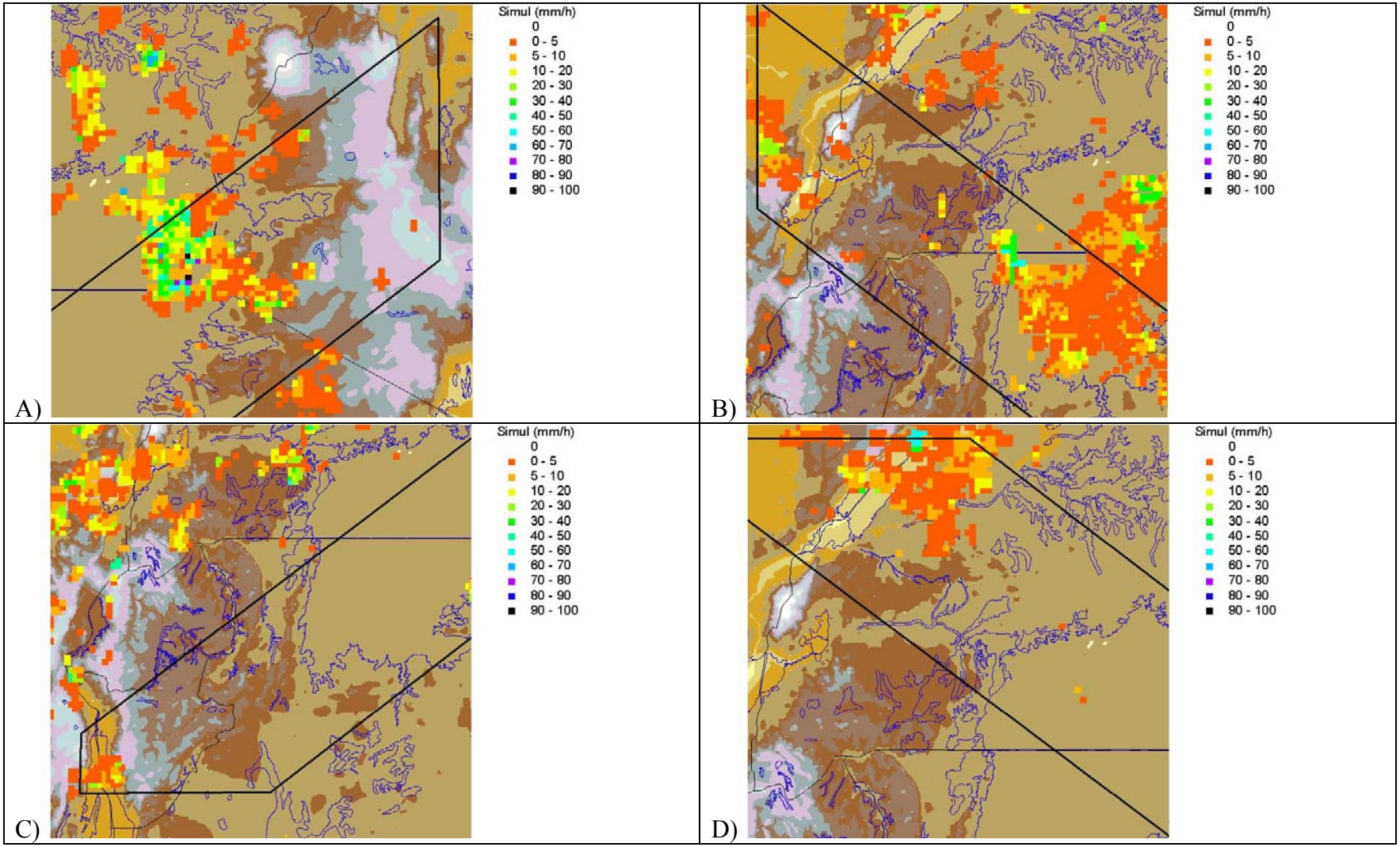


Figure 8.22 Realizations of single TRMM passes by indicator kriging with F-extension. A) Jan 09, 1998 22:12 GMT; B) May 3, 1998 23:26 GMT; C) May 04, 1998 10:50 GMT. D) Observed and model spatial correlation. July 17, 1998 08:40 GMT.

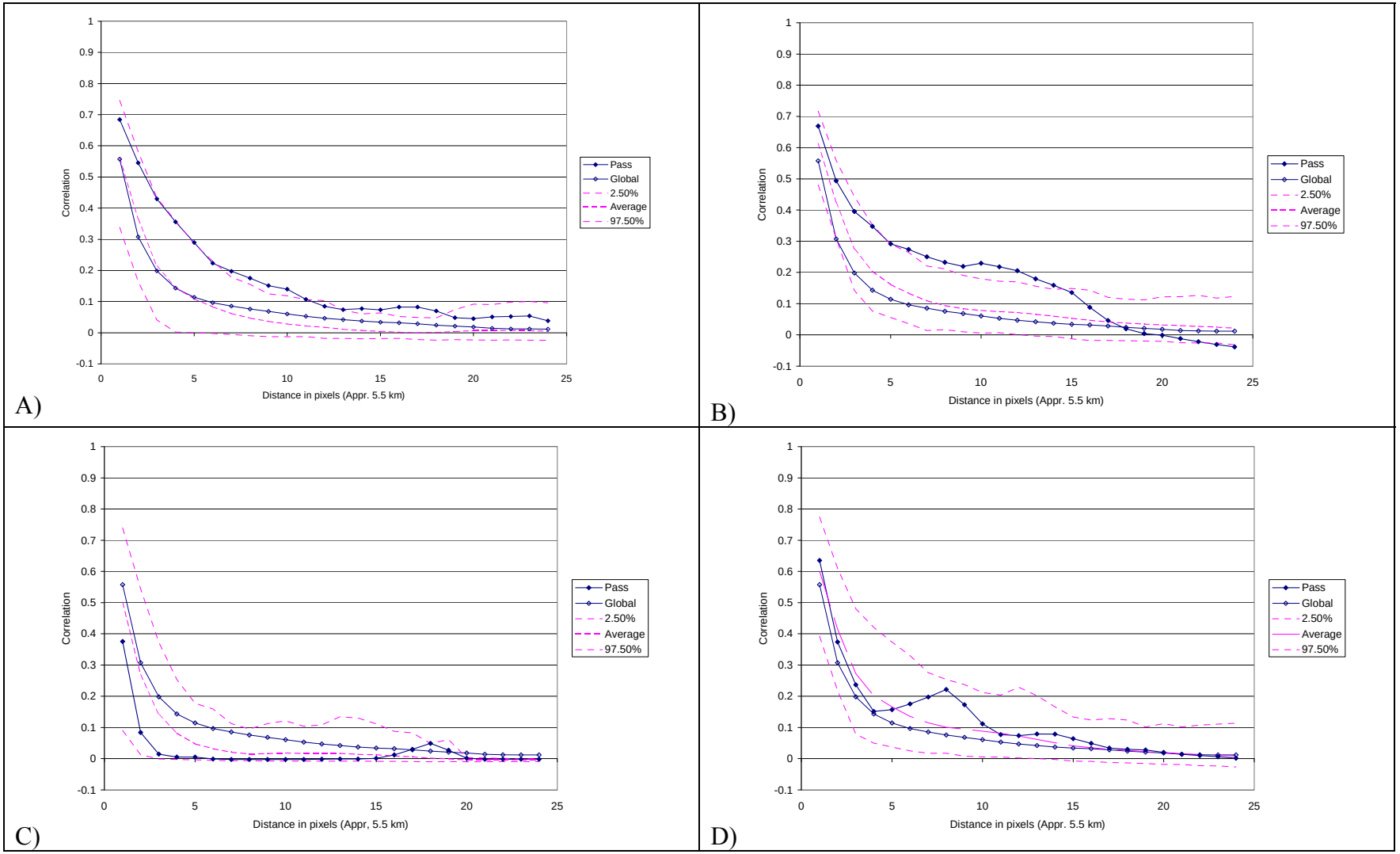


Figure 8.23 Spatial correlation of single TRMM passes by indicator kriging with F-extension. A) Jan 09, 1998 22:12 GMT; B) May 3, 1998 23:26 GMT; C) May 04, 1998 10:50 GMT. D) Observed and model spatial correlation. July 17, 1998 08:40 GMT.

An advantage of this approach over the use of a common 3x3 MAP is that it still takes into consideration the prior-cdf of each of the nine pixels, better simulating areas where precipitation shows high precipitation variability (i.e., the storm leading edge). A possible disadvantage is that areas of intense precipitation may be generated at a higher than the actual frequency.

The realizations reported in Figure 8.22 denote rain areas more contiguous and regular than those produced by normal indicator kriging. This results in increasing the precipitation spatial correlation (Figure 8.23). The short distance spatial correlation is much closer to the observed and, during large storms, can now exceed the average observed correlation. The medium and long range spatial correlation of the model remains smaller than the correlation sometimes observed during large storms, but this is due to the fact that only one average correlation function is used for both large and small storms.

Table 8.2 Observed and simulated average rain (mm h^{-1}) for the empirical distribution method.

	TRMM SR	Uncorrelated	Simple Kriging	Simple Kriging + F-Extension
Jan 09, 1998 22:12	0.96	0.47	0.51	0.42
May 3, 1998 23:26	1.02	0.95	1.03	0.93
May 4, 1998 10:50	0.11	0.13	0.12	0.11
July 17, 1998 08:40	0.50	0.45	0.46	0.42
Total	2.59	2.0	2.11	1.88

Table 8.2 shows that kriging methods can introduce bias in the estimates that, on average, is around 5-6% of the uncorrelated estimate. These biases are caused mainly by the variability of the prior cdf , the linear interpolation of the thresholds, and the extension procedure. It would be possible to introduce some filtering of the lower rain rates or other measures to correct it. However, it is difficult to do these adjustments in a

consistent manner, especially in view of the fact that the spatially correlated simulation is done only for a small area of the Lake Victoria basin. A correction that is right for the EKJ area may not be right for the mountains or the southern shore of the lake. Even inside the EKJ area there is variability due to the geographical differences. Consequently, there is the need to design a more standardized way to proceed.

The method proposed in this study is to correct the precipitation of the spatially correlated model to match the monthly average precipitation of the corresponding uncorrelated methods on a pixel-by-pixel basis. The correction is done by subtracting the bias from the daily precipitation to minimize the effect on the areas of light rain. The use of the uncorrelated models as benchmark is warranted by the fact that they have been calibrated and verified over the entire watershed and they could be easily updated in an operational setting.

8.6.1.3 Performances of the Temporally Uncorrelated Model

The temporally uncorrelated empirical-distribution model has been applied with the indicator residual kriging with 9-pixel F-extension and bias correction to estimate precipitation over the EKJ area according to the test format described at the beginning of the Chapter. Due to the decrease in computation time granted by the F-extension, the number of conditioning pixels has been increased to 15.

The model reproduces the gage spatial correlation very well with a gain of around 20 percentage points at the short distances (Figure 8.24). At very long distances, model correlation is a bit lower than the gage correlation, probably due to the failure of the average instantaneous-precipitation covariance to capture the long-range correlation of extensive storms.

The behavior of the average of the ensemble is similar to that of the spatially uncorrelated case (Figure 8.8 and Figure 8.25). Note that correlation and MAE at the

daily and monthly levels do not seem to be affected by the bias correction, which adjusts the monthly bias by increasing or decreasing the daily precipitation.

On the other hand, the decrease in 95% compliance with the number of gages declines dramatically to only 10, 3, and 8 percentage points at the daily, dekadal and monthly resolutions respectively.

The improvement in both spatial correlation and model performances show that the approach taken for modeling a spatially correlated random field is successful.

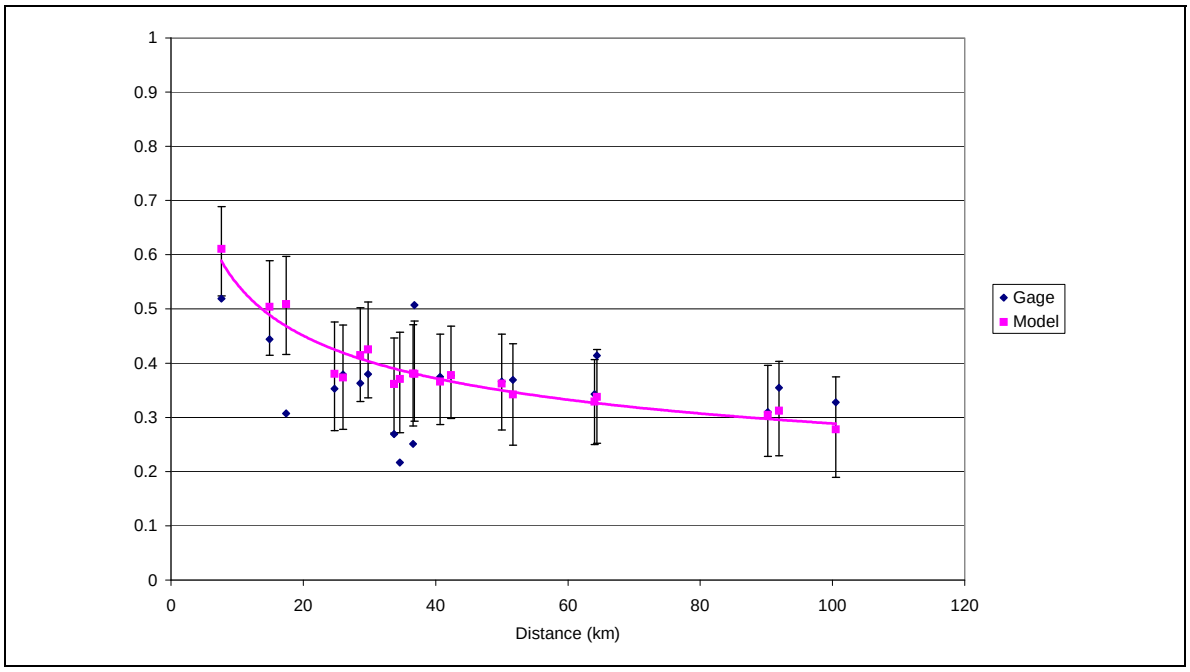


Figure 8.24 Correlation in daily precipitation for the EKJ area in 1996-1997. Temporally-uncorrelated empirical distributions with indicator residual kriging, 9-pixel F-extension, and bias correction.

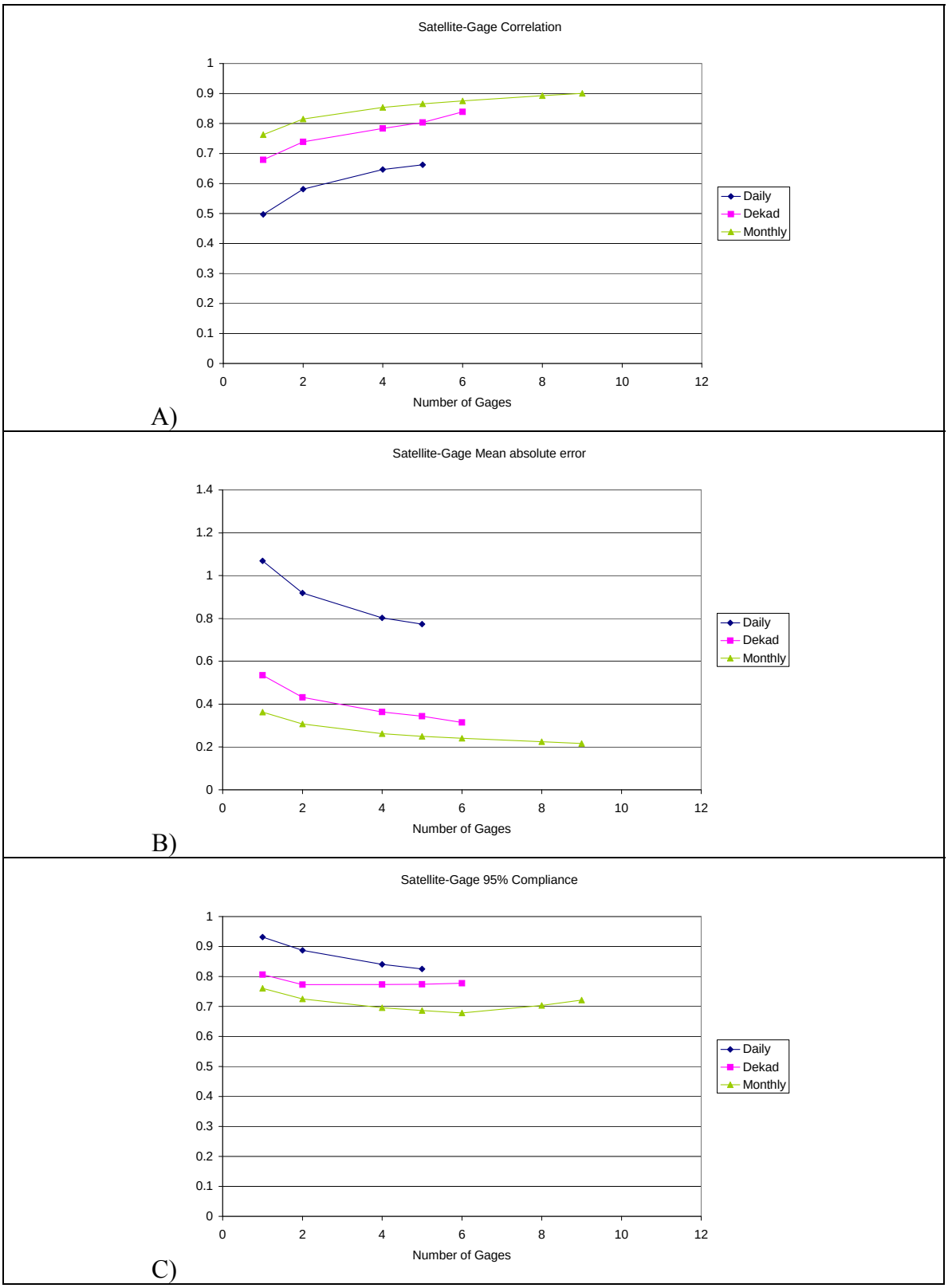


Figure 8.25 Multipixel performances in the EKJ area for temporally-uncorrelated empirical distributions with indicator residual kriging, 9-pixel F-extension, and bias correction.

8.6.1.4 Performance of the Temporally Correlated Model

The temporally correlated empirical-distribution model has been applied with the indicator residual kriging, 9-pixel F-extension, and bias correction to estimate precipitation over the EKJ area according to the test format described at the beginning of the Chapter. Thanks to the decrease in the computation time granted by the F-extension, the number of conditioning pixels has been increased to 15. It is noted that in the empirical-distribution model, the choice of the empirical distribution to use depends on the same-pixel precipitation at *t-1*. Therefore, at time *t*, the set of prior cdfs for one realization is chosen based on the precipitation field at time *t-1*, the IR, and storm stage at time *t*. The kriging procedure is then applied using this set of prior cdfs as outlined in section 8.6.1.2. Also, simulation rain below 0.5 mm h^{-1} has been discarded assuming that it has been generated by numerical noise.

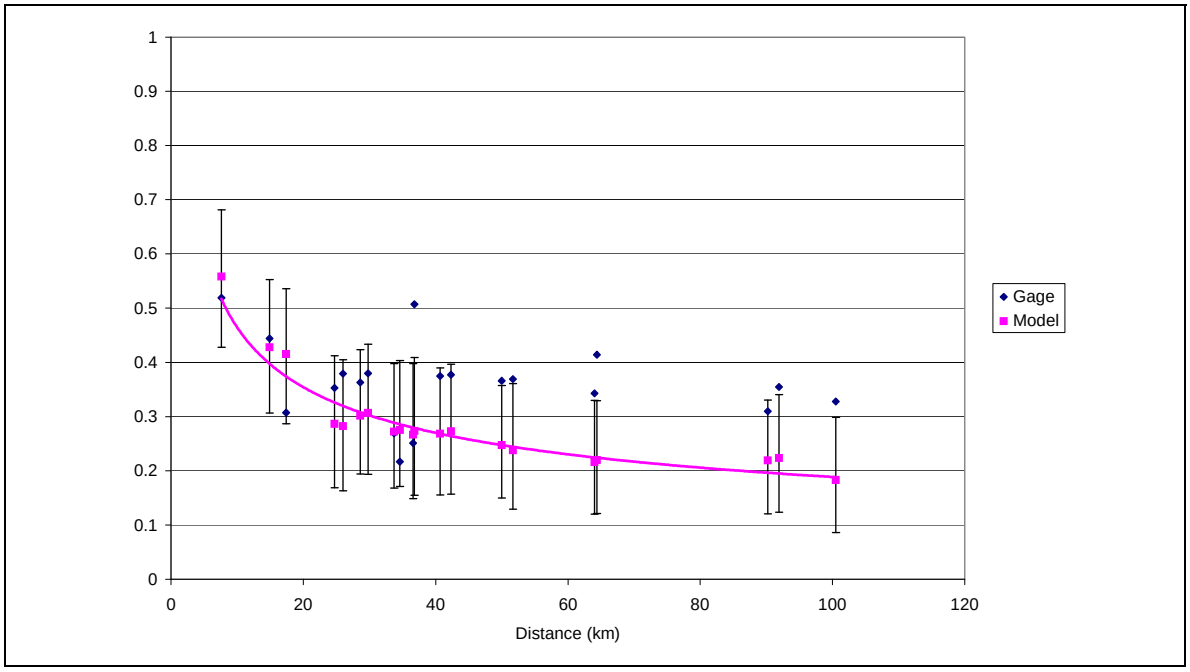


Figure 8.26 Correlation in daily precipitation for the EKJ area in 1996-1997. Temporally-correlated empirical distributions with indicator residual kriging, 9-pixel F-extension, and bias correction.

The modified SISIM algorithm increases the precipitation spatial correlation also for the temporally correlated model, performing well for distances below 20 km and acceptably for distances below 40 km. At higher distances, the low spatial correlation of the instantaneous precipitation is not able to smooth the randomness introduced by the same pixel temporal correlation (Figure 8.26).

Figure 8.27 shows that the correlation between gages and the ensemble average increases with the number of gages considered. However, as anticipated at the beginning of the Chapter, at the daily level the increase is lower for small areas than it is for large. This is due to the enhancement caused by averaging over larger areas. At the decadal and monthly levels, on the other hand, this effect is almost completely mitigated. Actually, at the monthly level, the trend seems reversed, with the correlation increasing much more over small areas than over large areas. The number of data is insufficient to conclude if this counter-intuitive result is real or it is simply an effect of a particular combination of gages.

Figure 8.28 shows the "dual" of this relation for the mean absolute error: MAE decreases with the number of gages, but it does so at lower pace when considering neighboring pixels than when considering more sparse combinations of pixels.

The Figure 8.29 charts demonstrate that the decline in compliance rates with the number of gages is not really due to the number of pixels included, but to the increase in the area covered by the combinations. This can be seen by the fact that, when separated according to the size of the gage combination, compliances are distributed along nearly constant straight lines at levels decreasing with the size of the combination. The small decrease in compliance within the same class of combination sizes is due to the effect of specific combinations and to the marginal increase in the average size of combinations within each class (Figure 8.30).

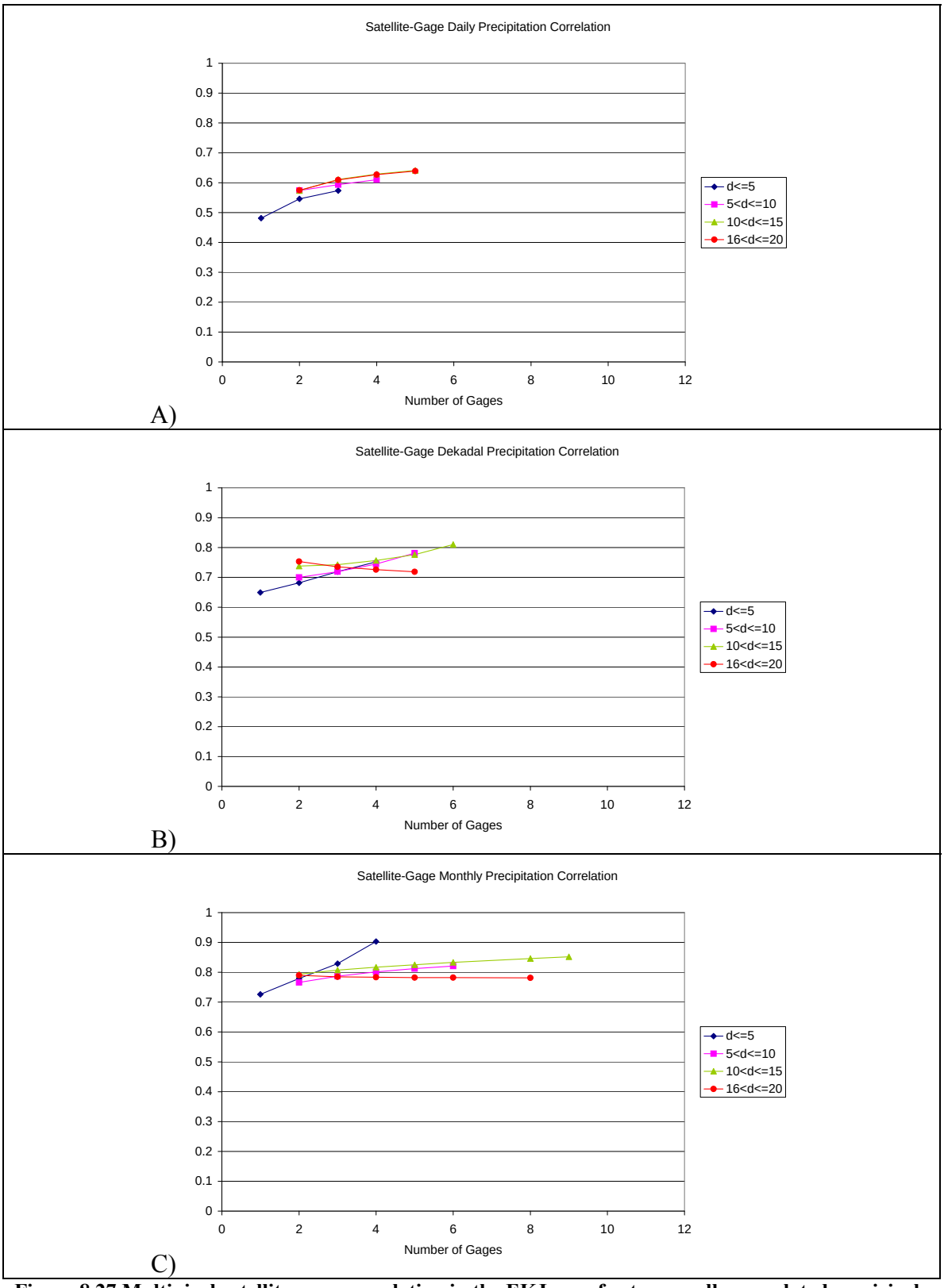


Figure 8.27 Multipixel satellite-gage correlation in the EKJ area for temporally-correlated empirical distributions with indicator residual kriging, 9-pixel F-extension, and bias correction.

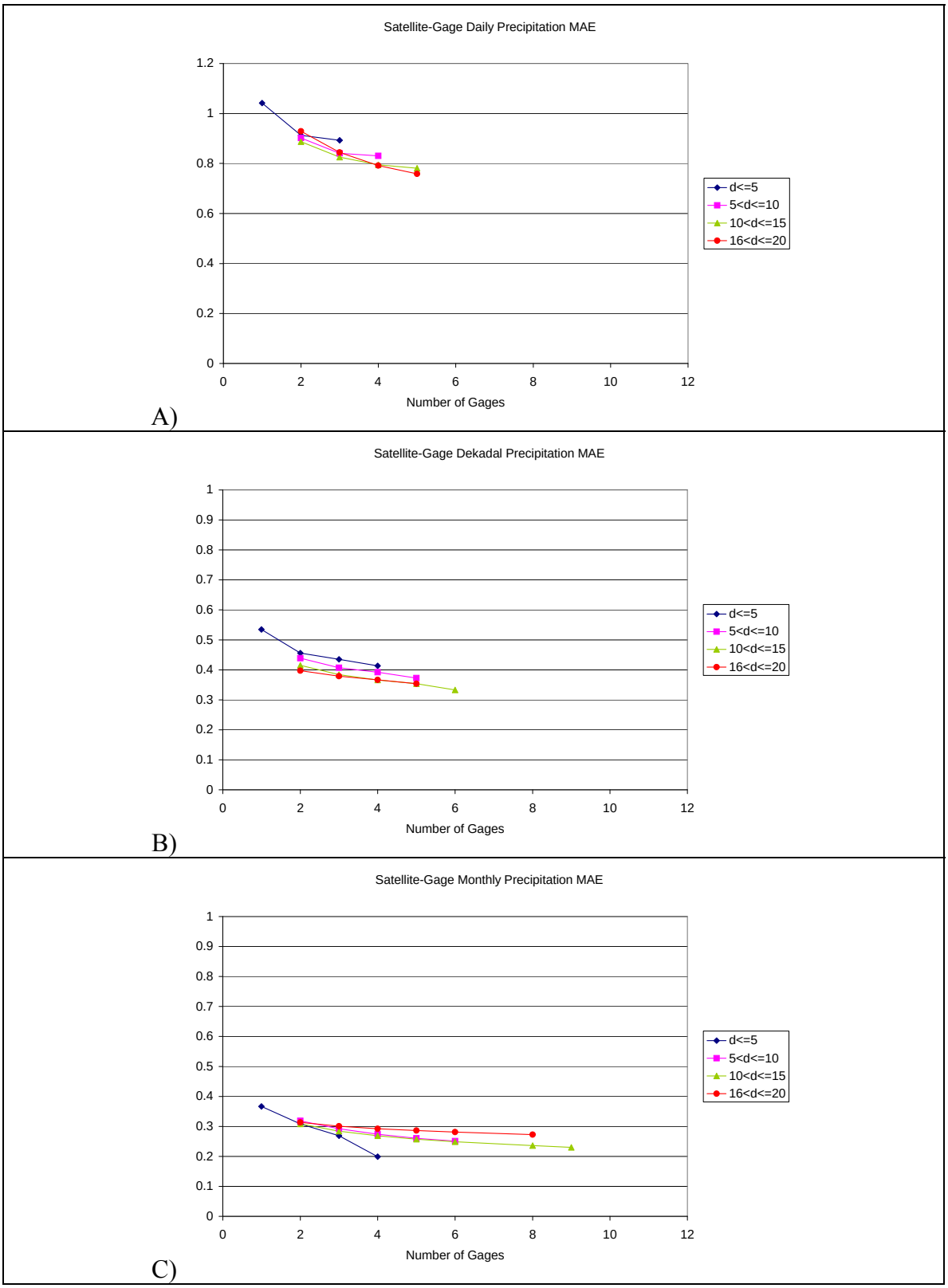


Figure 8.28 Multipixel satellite-gage Mean Absolute Error in the EKJ area for temporally-correlated empirical distributions with indicator residual kriging, 9-pixel F-extension, and bias correction.

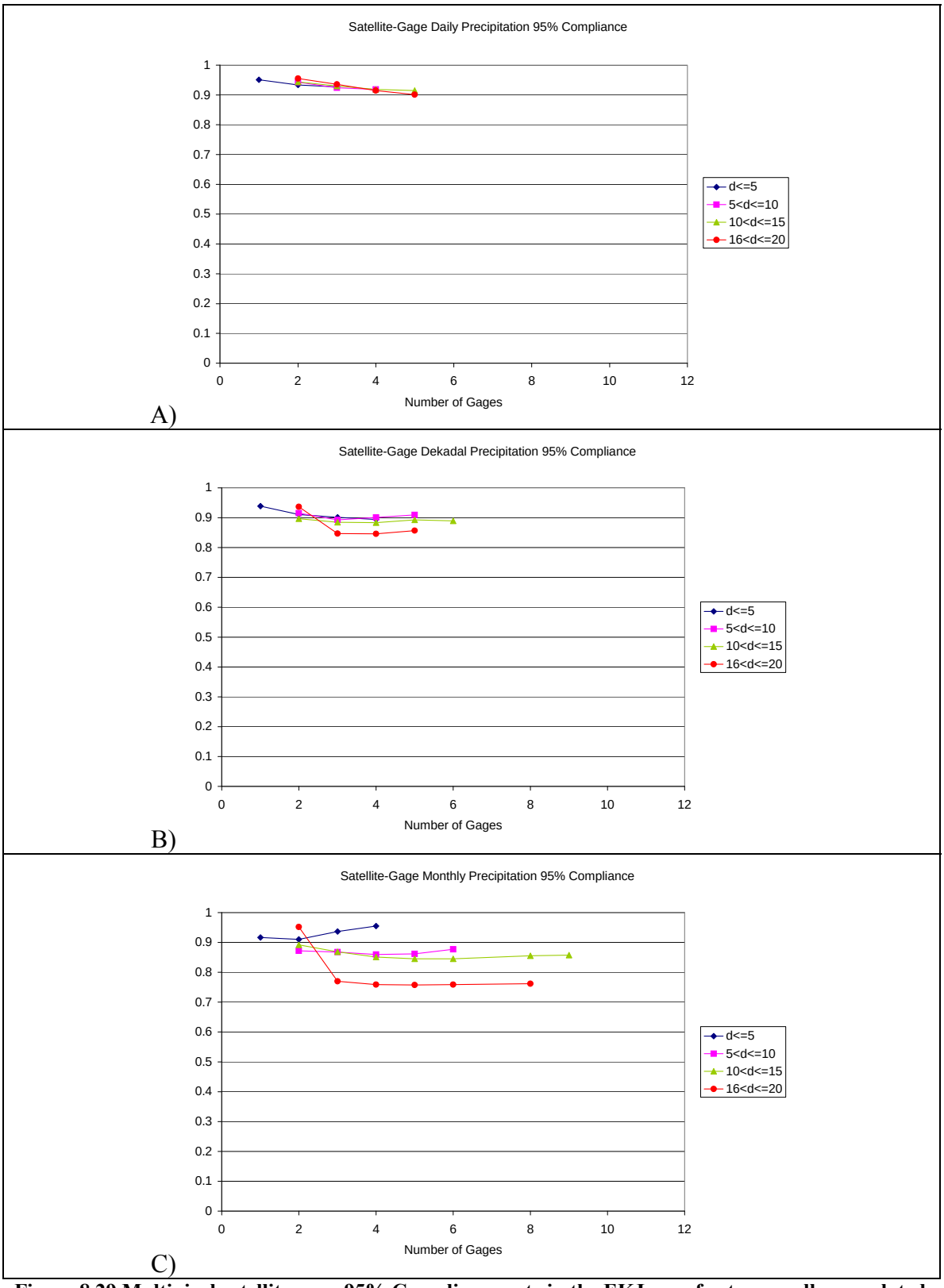


Figure 8.29 Multipixel satellite-gage 95% Compliance rate in the EKJ area for temporally-correlated empirical distributions with indicator residual kriging, 9-pixel F-extension, and bias correction.

The lower compliance rate for sparser combinations is instead caused by the poor replication of the spatial correlation at medium to large distances, which reduces the variability of precipitation over the combinations of pixels more than it is warranted.

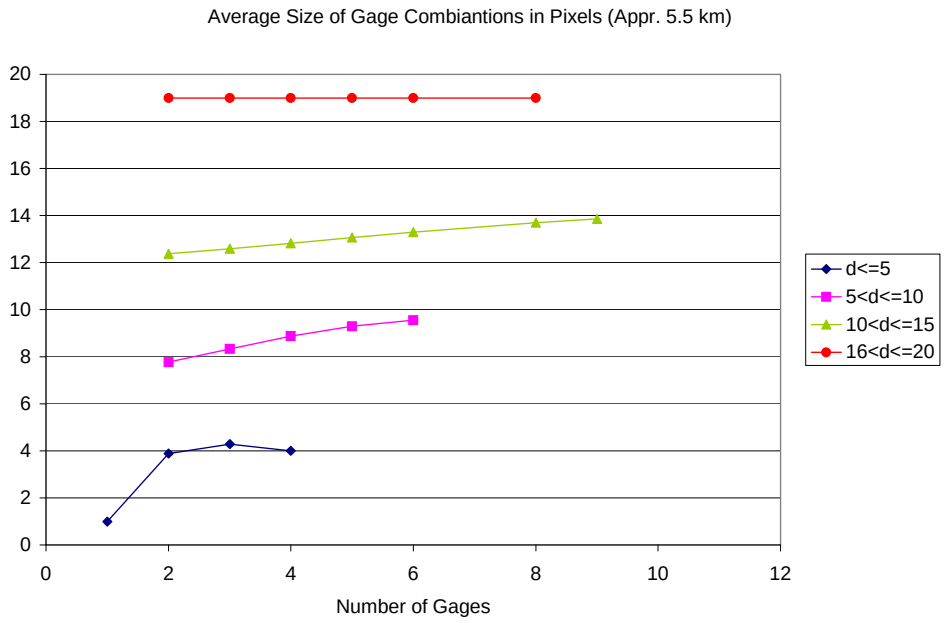


Figure 8.30 Average size of the smallest square entirely containing the combination of gages as a function of the number of gages and size class.

If the compliance rate is not a function of the number of elements, but it is a function of their distance, then the reliability of the precipitation estimates for squares of a certain size can be approximated by the reliability of the combinations of gages contained within such squares. One could argue that the average of all possible combinations contained within a square of size d is an index of its reliability. Here, a more conservative approach is taken using as reliability the minimum average reliability for combinations of the same number of elements within the square box (Table 8.3). The reliability is satisfactory at the daily and decadal resolutions for all sizes of practical

interest. At the monthly level, the reliability decreases substantially for larger areas. Part of this decline could be eliminated by increasing the pool of contemporaneous geostationary and TRMM data used for deriving the precipitation distributions. As pointed out in Chapter 7, the monthly compliance rate increased when deriving the precipitation distribution using data from only one or two consecutive months, instead of three months, as it was deemed originally necessary.

Furthermore, a better modeling of the spatial autocorrelation for large storms and possibly a reduction of the model spatial resolution to 2x2 or 3x3 would also contribute to rectify this problem (by smoothing the prior cdf and decreasing the effect of the temporal correlation on the precipitation spatial correlation).

Table 8.3 95% Compliance of the temporally correlated, spatially correlated empirical distribution model.

Size of square	Daily	Dekadal	Monthly
<=0.25°	0.93	0.89	0.91
<=0.50°	0.92	0.89	0.86
<=0.75°	0.92	0.89	0.85
<=1.00°	0.91	0.88	0.84

On the other hand, Figure 8.31 shows that averaging the precipitation over a number of gages leads to a rapid decrease of the estimate uncertainty. The figure shows that only five gages are sufficient to reduce the uncertainty by almost 50%. The uncertainty continues to decline for larger numbers of gages, albeit at a lower rate (Figure 8.31 C)). The uncertainty reduction increases with the size of the combination likely because the spatial correlation is lower at medium-large distances.

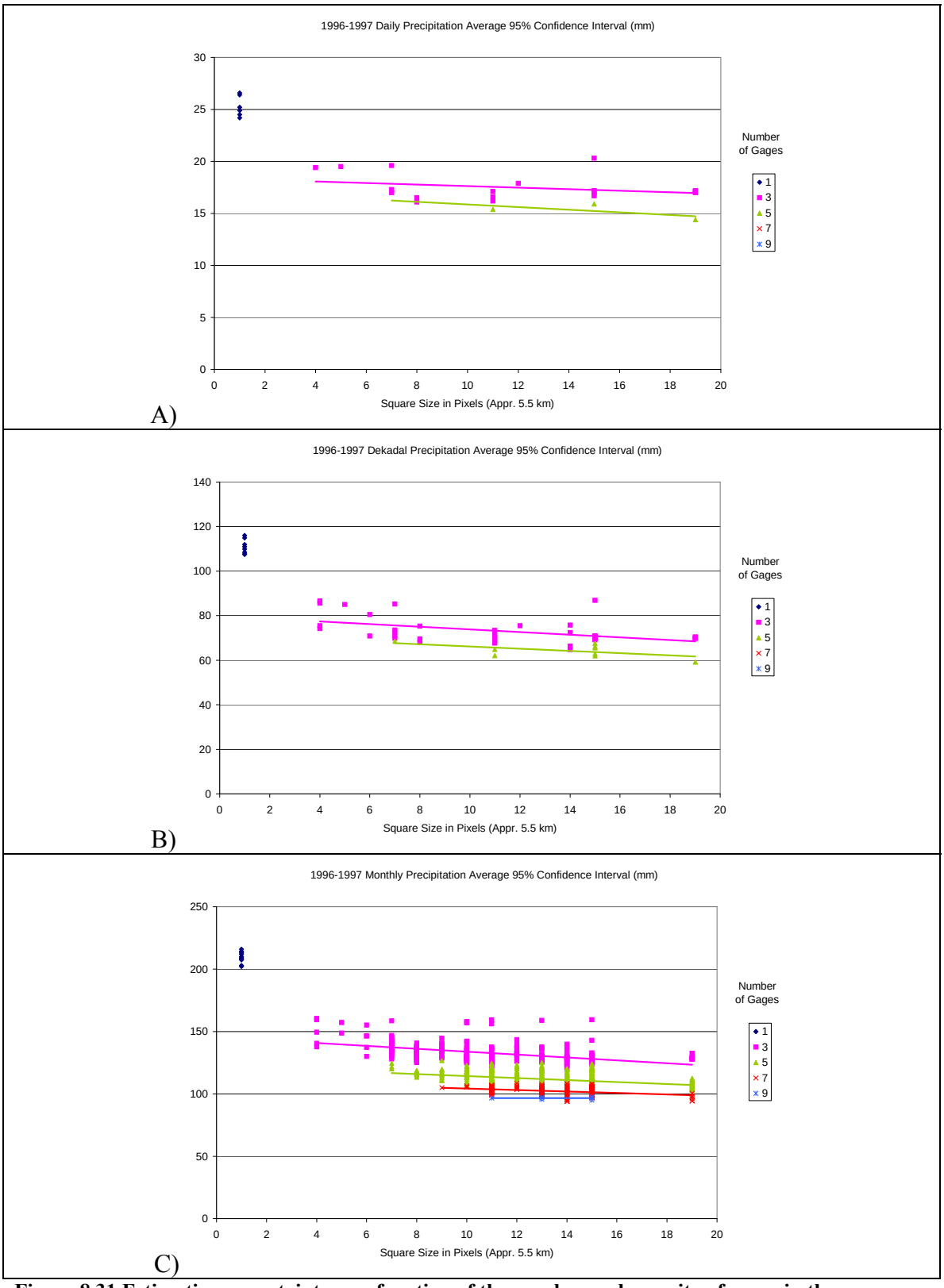


Figure 8.31 Estimation uncertainty as a function of the number and sparsity of gages in the combination. Temporally correlated empirical distributions with indicator residual kriging, 9-pixel F-extension, and bias correction

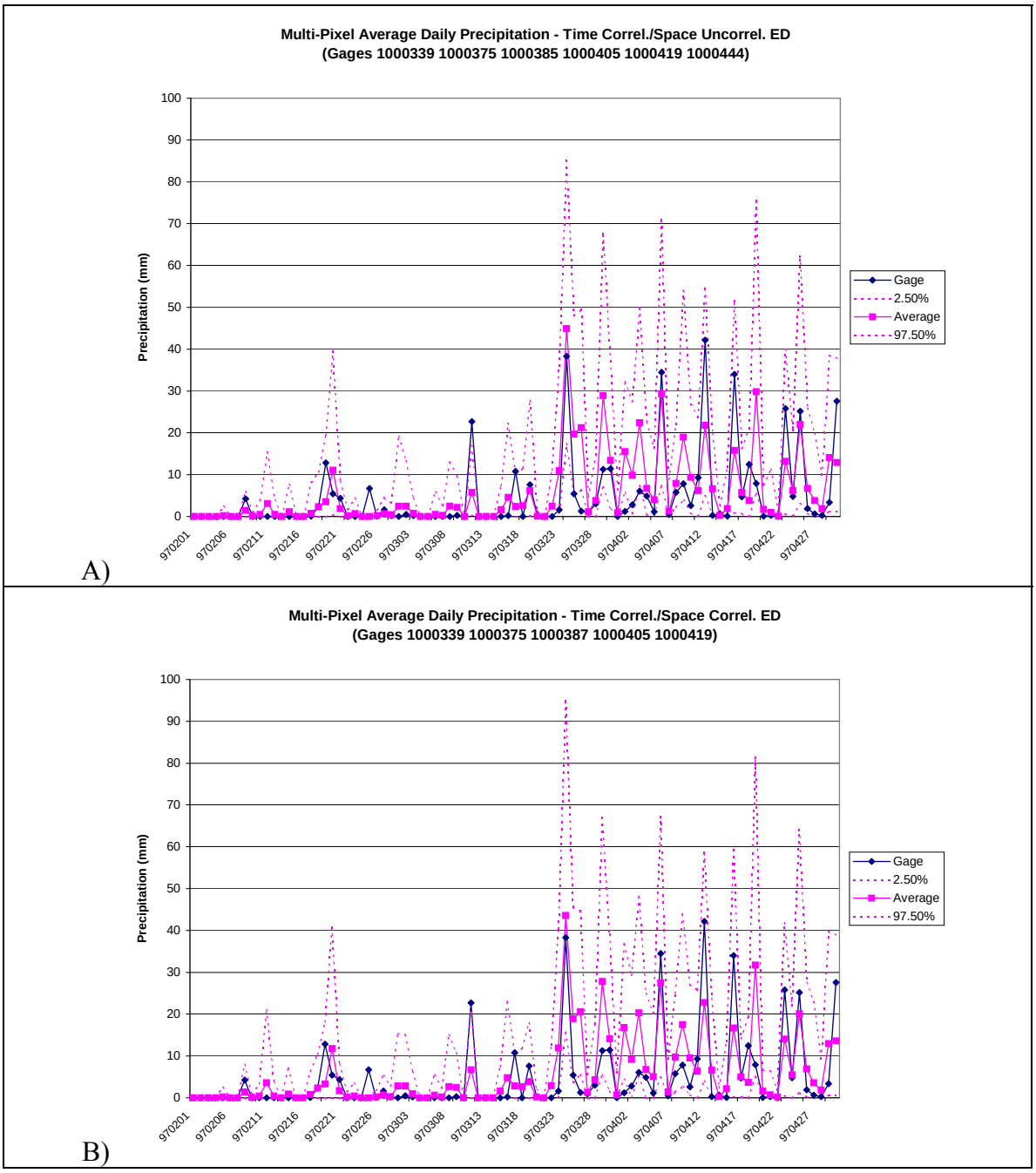
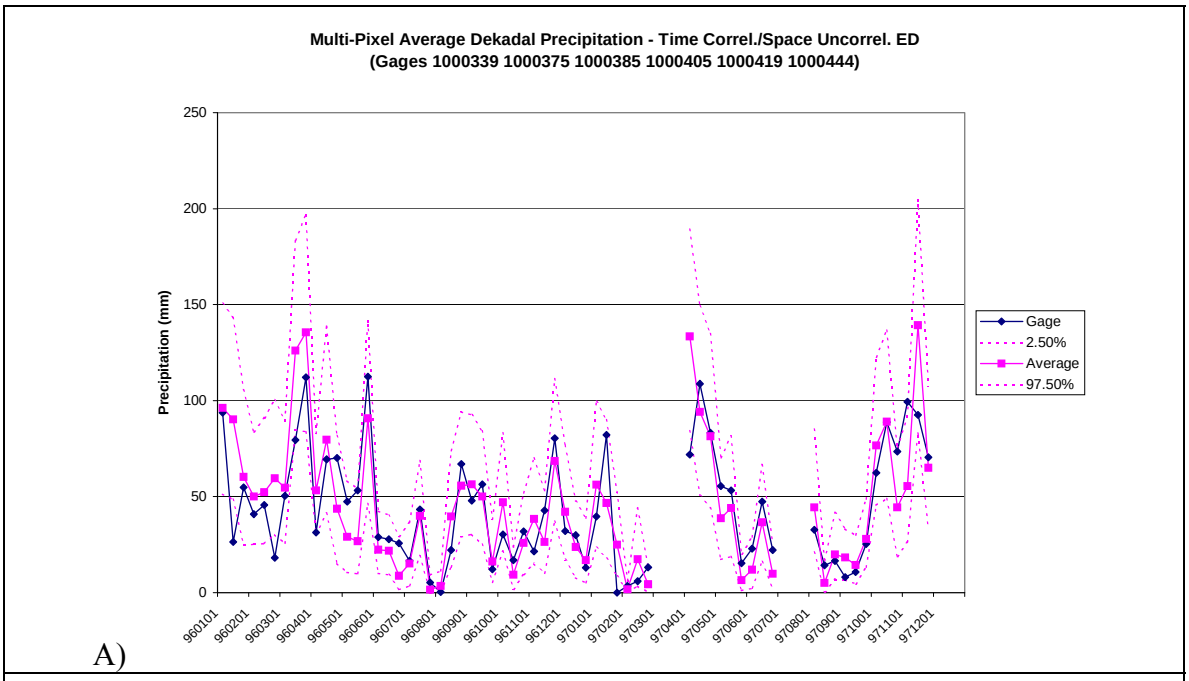


Figure 8.32 Daily precipitation at a combination of five gages in the EKJ area. Temporally correlated empirical distributions A) Spatially uncorrelated; B) Indicator residual kriging, 9-pixel F-extension, and bias correction.

Figure 8.32 - Figure 8.34 show that the estimation methodology is able to reproduce the observed average precipitation over a combinations of several gages, especially at the decadal and monthly resolutions.



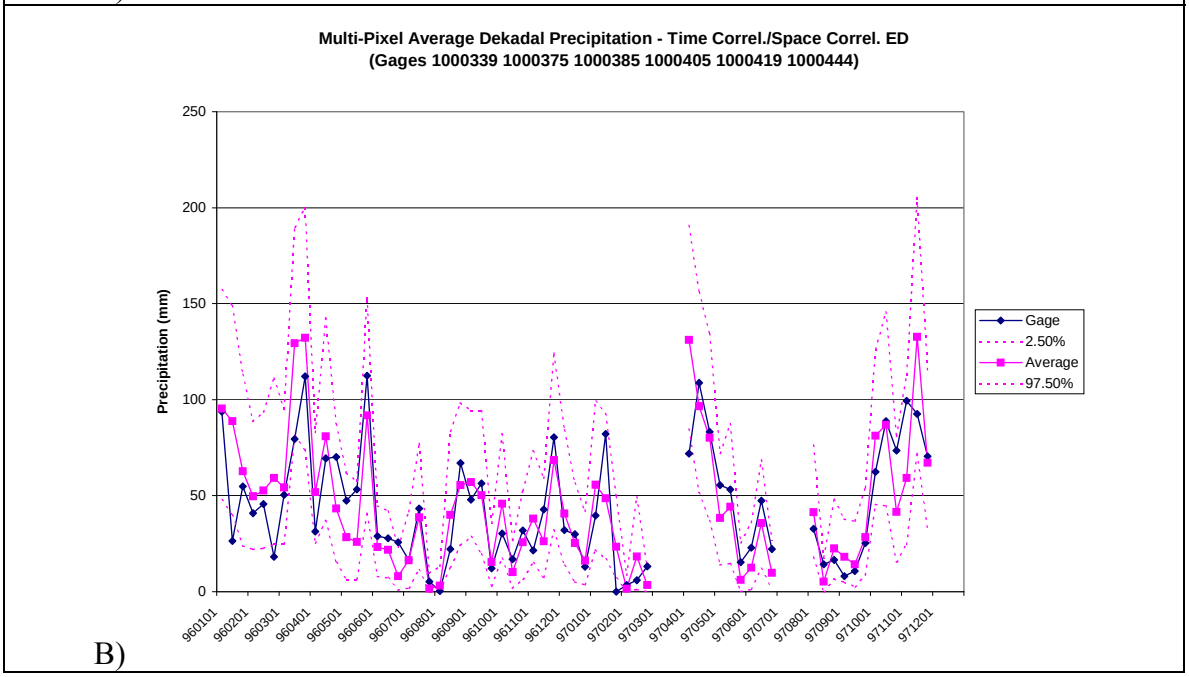
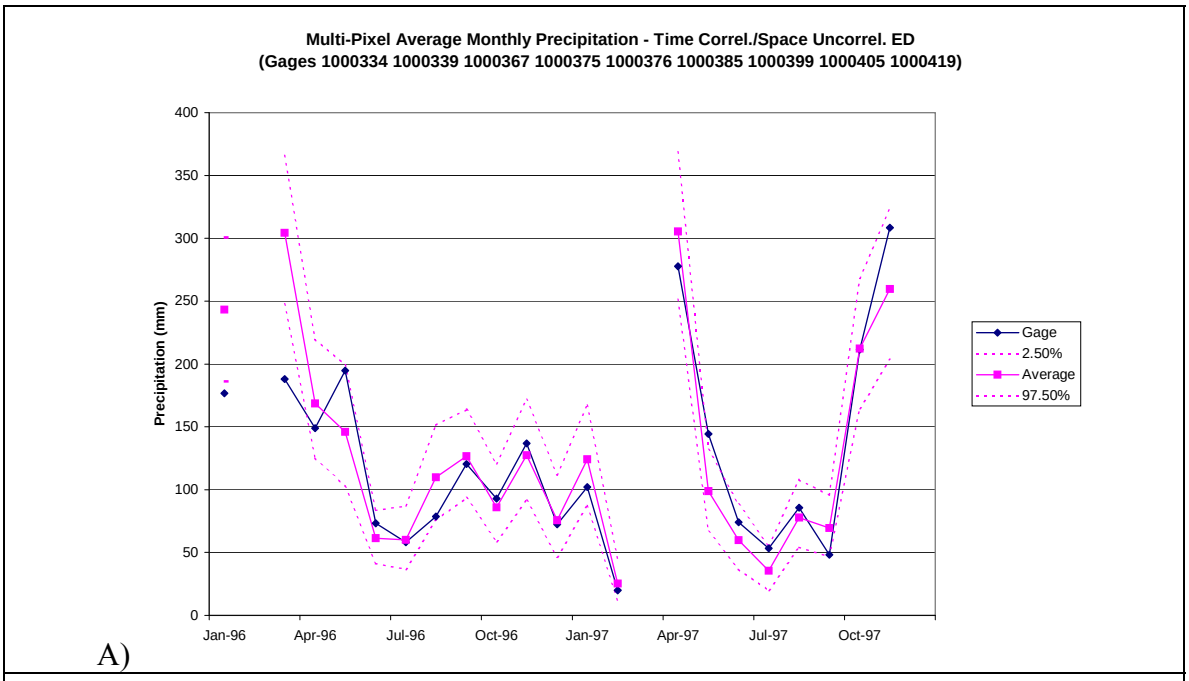
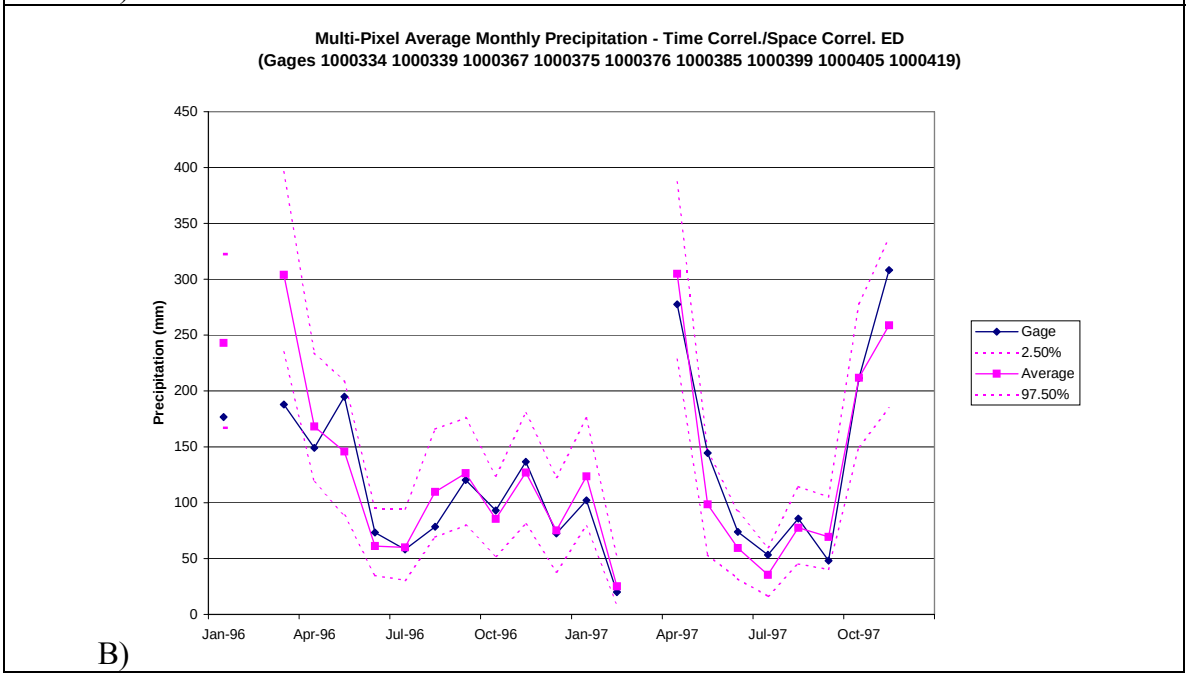


Figure 8.33 Dekadal precipitation at a combination of six gages in the EKJ area. Temporally correlated empirical distributions A) Spatially uncorrelated; B)Indicator residual kriging, 9-pixel F-extension, and bias correction.

The spatially-correlated and spatially-uncorrelated methods estimate gage precipitation in almost the same way, but the spatially uncorrelated method generate narrower variability bands, that miss several data points more than the spatially correlated estimator.



A)



B)

Figure 8.34 Monthly precipitation at a combination of nine gages in the EKJ area. Temporally correlated empirical distributions A) Spatially uncorrelated; B)Indicator residual kriging, 9-pixel F-extension, and bias correction.

8.6.2 Analytical Distributions, Sequential Indicator/Gaussian Simulation

As introduced in Section 8.5.5, a two-step process is adopted to realize spatially-correlated precipitation fields when using the analytical distributions. In the first step, the rainy pixel distribution is established with the same SISIM approach described in Section 8.6.1, but limited only to the 0.0 threshold (equivalent to 0.0 mm h^{-1}). In the second step, the precipitation intensity is simulated using a SGSIM approach. The approach follows that proposed by Barancourt et al. (1992) and Pardo-Iguzquiza et al. (2006). There are, however, some differences with the work of these investigators. First, the precipitation distribution is modeled as lognormal, as suggested by Bell (1987), and not normal as in Barancourt et al. (1992). Pardo-Iguzquiza et al. (2006), on the other hand, use the normal-score transform to normalize their data. Second, the procedure does not use "hard" data to condition the simulation, but only soft data (a-priori $F_R(0,x)$, $\mu_{LNR}(x)$, and $\sigma_{LNR}(x)$). Third, the soft conditioning makes the simulation non-stationary.

The simulation of the rain/no-rain distribution is done as in Section 8.6.1 using a spatially-varying a-priori distribution of $F_R(0,x)$ and the stationary indicator residual covariance function determined in Section 8.6.1.1.

On the other hand, the assumption of covariance stationarity in simulating the rain intensity would imply that the precipitation estimates have a spatially varying mean $\mu_{LNR}(x)$ but constant variance $\sigma_{LNR}(x)$. The analytical distributions derived in Chapter 6, however, are characterized by very different variances depending on the satellite information. Thus the assumption that $\sigma_{LNR}(x)$ is constant in space would be an oversimplification.

A possible approach is normalizing the logarithm of the positive TRMM rain rates through the local rain $\mu_{LNR}(x)$ and $\sigma_{LNR}(x)$ as follows:

$$Y(\mathbf{x}) = \frac{\ln(Z(x)) - \mu_{LNR}(\mathbf{x})}{\sigma_{LNR}(\mathbf{x})} \tag{8.26}$$

By construction, *Y(**x**)* is an identically distributed multivariate normal distribution, with zero mean and unit variance. Thus it would be possible to use the classical kriging approach to unconditionally simulate *Y(x)* and then back-transform the simulation results to *Z(x)* using the inverse of 8.26 and the logarithmic anti-transformation. The covariance function of the random field *Y* needs to be estimated from the TRMM data transformed according to 8.26.

An alternative approach is proposed by Fiorucci et al. (2001), who modeled directly the *ln(z(x))* variable and set the covariance as the product of the local variance for a stationary correlation coefficient:

$$\mathrm{cov}_{LNR}(\mathbf{x},\mathbf{y}) = \sigma_{LNR}(\mathbf{x})\sigma_{LNR}(\mathbf{y})\rho_{LNR}(\mathbf{x},\mathbf{y}) = \sigma_{LNR}(\mathbf{x})\sigma_{LNR}(\mathbf{y})\rho_{LNR}(|\mathbf{x}-\mathbf{y}|) \tag{8.27}$$

The two approaches are equivalent, but the approach by Fiorucci et al. (2001) features one data transformation less, keeping closer to the original data, and is preferred here. The equations used for estimating the mean of the conditional lognormal distribution based on the precipitation previously simulated at M neighboring points is derived adapting equations 8.15-8.17 to the non-stationary case:

$$\begin{cases} E\{Z(x_j) \mid z(x_k), K=1,..,M\} = [z(x_j)]_{SK} = m_{LNR}(x_j) + \sum_{k=1}^{M} a_k [z(x_k) - m_{LNR}(x_k)] \\ \sum_{k=1}^{M} a_k C_{LNR}(x_k, x_i) = C_{LNR}(x_j, x_i), \qquad i1, M \\ \mathrm{C_{LNR}}(\mathrm{x_j}, \mathrm{x_k}) = \sigma_{\mathrm{LNR}}(\mathrm{x_j})\sigma_{\mathrm{LNR}}(\mathrm{x_k})\rho_{\mathrm{LNR}}(|\mathrm{x_j} - \mathrm{x_k}|) \end{cases} \tag{8.28}$$

Where $m_{LNR}(x_k)$ = priori mean of the *ln(r(x_k))* based on satellite information

$\sigma_{LNR}(x_k)$ = priori standard deviation of the residuals *(ln(r(x)-m_{LNR}(x))* based on satellite information

$C_{LNR}(x_k,x_j)$ = covariance of the residuals *(ln(r(x)-m_{LNR}(x))* between points x_k and x_j

$\rho_{LNR}(x_k,x_j)$ = correlation of the residuals *(ln(r(x))- m_{LNR}(x))* between points x_k and x_j

The (conditional) variance of the logarithm of precipitation intensity *Z(xj)* is identified with the simple kriging variance:

$$VAR\{Z(x_j) \mid z(x_k), k = 1,..,M\} = C_{LNR}(x_j, x_j) - \sum_{k=1}^{M} a_k C_{LNR}(x_j, x_k) \qquad (8.29)$$

8.6.2.1 Modeling of the Precipitation Residual Covariance

The covariance functions used for simulating the rain/no-rain distribution are the same derived in Section 8.6.1.1 for the threshold 0.0, so they will not be re-examined here. As mentioned before, the approach taken is to model $\rho_{LNR}(x_k,x_j)$ as a stationary function, and then derive the $C_{LNR}(x_k,x_j)$ according to 8.28. As in Section 8.6.1.1, the variogram of the precipitation logarithm residuals was modeled as the combination of two exponential models, one for the long range relations and one for the short range relations (equation 8.25). Figure 8.35 shows that, as expected from Sections 8.3 and 8.6.1.1, the correlation over lake pixels is stronger than over land pixels. However, probably because Lake Victoria is large, but limited, land pixels show a longer tail (i.e., a more persistent long range component) than lake pixels. Beyond 70 km, the lake's precipitation residual correlation is computed with an increasing number of land pixels, which have rain characteristics completely different from the lake pixels.

The resulting models reflect these conditions. The short distance component has a similar range in the two areas, but the range of the long distance component is larger over land. On the other hand, the short-range component is a larger fraction of the variogram for land pixels than it is for lake pixels. (It is noted that the variogram has been normalized to produce the correlation coefficient instead of the covariance.)

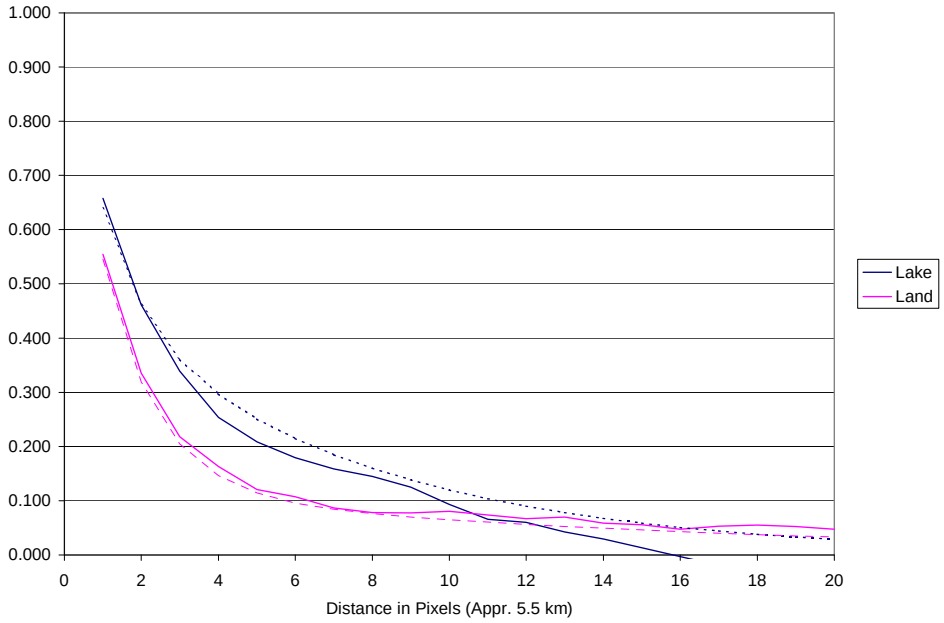


Figure 8.35 Sample (solid) and model (dashed) ln(r) residual correlation.

Table 8.4 The coefficients of the variogram models used in the SISIM simulation.

Threshold	n	c1	a1 (km)	c2	a2 (km)
Lake	0	0.596	19.1	0.404	81.7
Land	0	0.875	23.1	0.125	247.5

8.6.2.2 Comparison with single TRMM Passes

When the kriging is applied to all pixels, the mixed SISIM/SGSIM suffers of the same low spatial correlation seen in Section 8.6.1.2. Therefore, the algorithm was modified to emulate the 3x3 pixel spatial resolution by the F-extension introduced there. This is accomplished as follows:

1. Kriging is performed as in equation 8.22 for estimating the conditional probability of rain/no-rain at a point *x*, based on the neighboring indicator residuals.
2. The modifications of the probability of rain produced by kriging and the sampling probability used for determining the realization of the rain/no-rain random field in *x* are extended to the eight pixels adjacent to x, making the nine pixels behaving as a 3x3 area distribution.
3. The kriging is performed as in equation 8.28 for estimating the mean and variance of the precipitation intensity at a point *x*, based on the neighboring precipitation residuals.
4. The mean and variance modifications found in step 3 and the sampling probability used for determining the realization of the rain intensity in *x* are extended to the eight pixels adjacent to *x*, making the nine pixels behaving as a 3x3 area distribution.

The only relation between the two random fields is that the random path used to inspect the simulation grid is the same, so that the 3x3 areas are completely overlapping.

With the exception of the Jan 09, 1998 22:12 GMT pass, heavily dominated by "warm" orographic precipitation over the mountains, the percentage of rainy pixels of the temporally and spatially uncorrelated precipitation is relatively close to the observed one (Table 8.5). Average precipitation is close, but with higher variation. The normal kriging procedure increases both the number of rainy pixels and the average precipitation.

The extension of the kriging results to the nine pixels adjacent to the kriging pixel does not have a clear effect on the number of rainy pixels, although it slightly increases the average precipitation.

Consequently, the low-precipitation screening was increased from 0.5 to 0.7 mm h^{-1}, to bring the number of rainy pixels closer to the uncorrelated case. The bias correction procedure described in Section 8.6.1.2 was also adopted for simulating daily precipitation.

Table 8.5 Observed and simulated average precipitation for the analytical distribution method.

	TRMM SR	Uncorrelated	Simple Kriging	Simple Kriging + F-Extension
Rainy Pixels (%)				
Jan 09, 1998 22:12	16.9	6.5	7.6	7.2
May 3, 1998 23:26	15.9	16.6	17.6	17.6
May 4, 1998 10:50	2.5	1.7	1.8	1.8
July 17, 1998 08:40	7.3	5.6	5.8	5.8
Average	10.7	7.6	8.21	8.08
Average precipitation (mm)				
Jan 09, 1998 22:12	0.96	0.41	0.47	0.46
May 3, 1998 23:26	1.02	0.93	0.98	1.01
May 4, 1998 10:50	0.11	0.14	0.14	0.15
July 17, 1998 08:40	0.50	0.34	0.35	0.37
Total	2.59	1.81	1.93	2.00

The realizations in Figure 8.36 - Figure 8.39 feature extensive areas of relatively uniform precipitation similar to those present in the TRMM images, showing that the kriging procedure is successful. They also present contiguous areas of intense rain that are not common in the TRMM data and in the empirical distribution simulations. This may be partly due to the coarser IR/VIS/WV resolution of the analytical distributions, which may produce identical results for larger numbers of contiguous pixels. Another

factor is that the kriging on the mean and variance of the precipitation intensity is virtually unconstrained, since the filtering of the precipitation exceeding 1.25 times the a-priori maximum precipitation intensity is executed only on the final results of the simulation. However, the frequency of these intense rain patches is not too frequent, given that the histograms associated with the simulation do not feature particularly high 97.5% percentiles. On the contrary, on both January 09 and July 17, the 97.5% levels for several stages are lower than the corresponding values produced by the empirical distributions. This is not as much due to kriging as it is due to the characteristics of the analytical distributions. The higher coarseness of these distributions may result in a lower average precipitation rate. Furthermore, the annual precipitation distribution is used in place of the tri-monthly distributions if these have too coarse IR/VIS/WV resolution.

In addition, the kriging of the analytical distributions produces a lower spatial correlation than that of the empirical distributions, probably because the analytical distributions are noisier than the empirical ones (see Figure 8.7). The difference is however limited to a few percentage points.

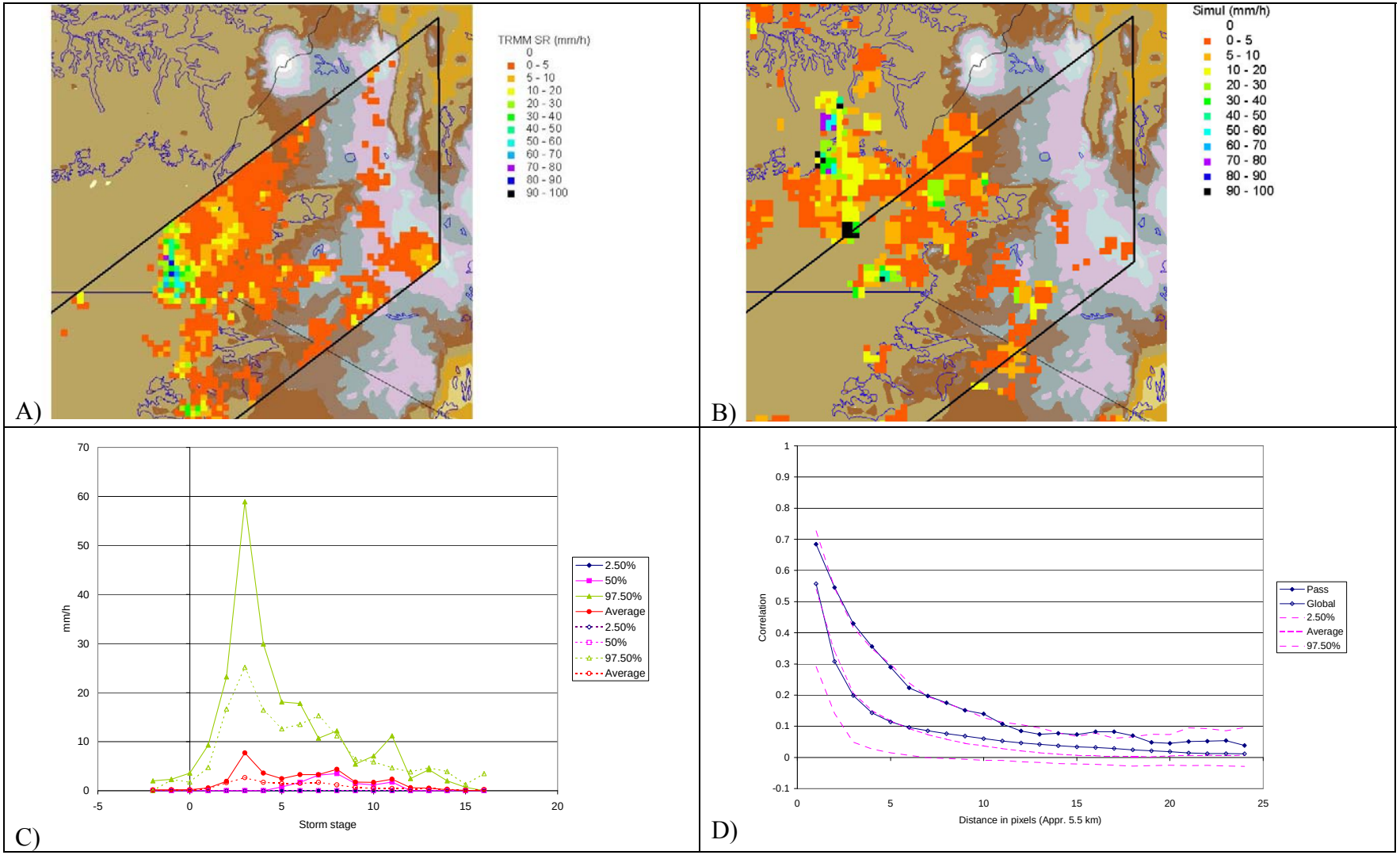


Figure 8.36 Simulation of single TRMM passes. A) Observed TRMM SR map; B) One realization of the TRMM rain map; C) Observed and model precipitation distribution. D) Observed and model spatial correlation. Jan 09, 1998 22:12 GMT

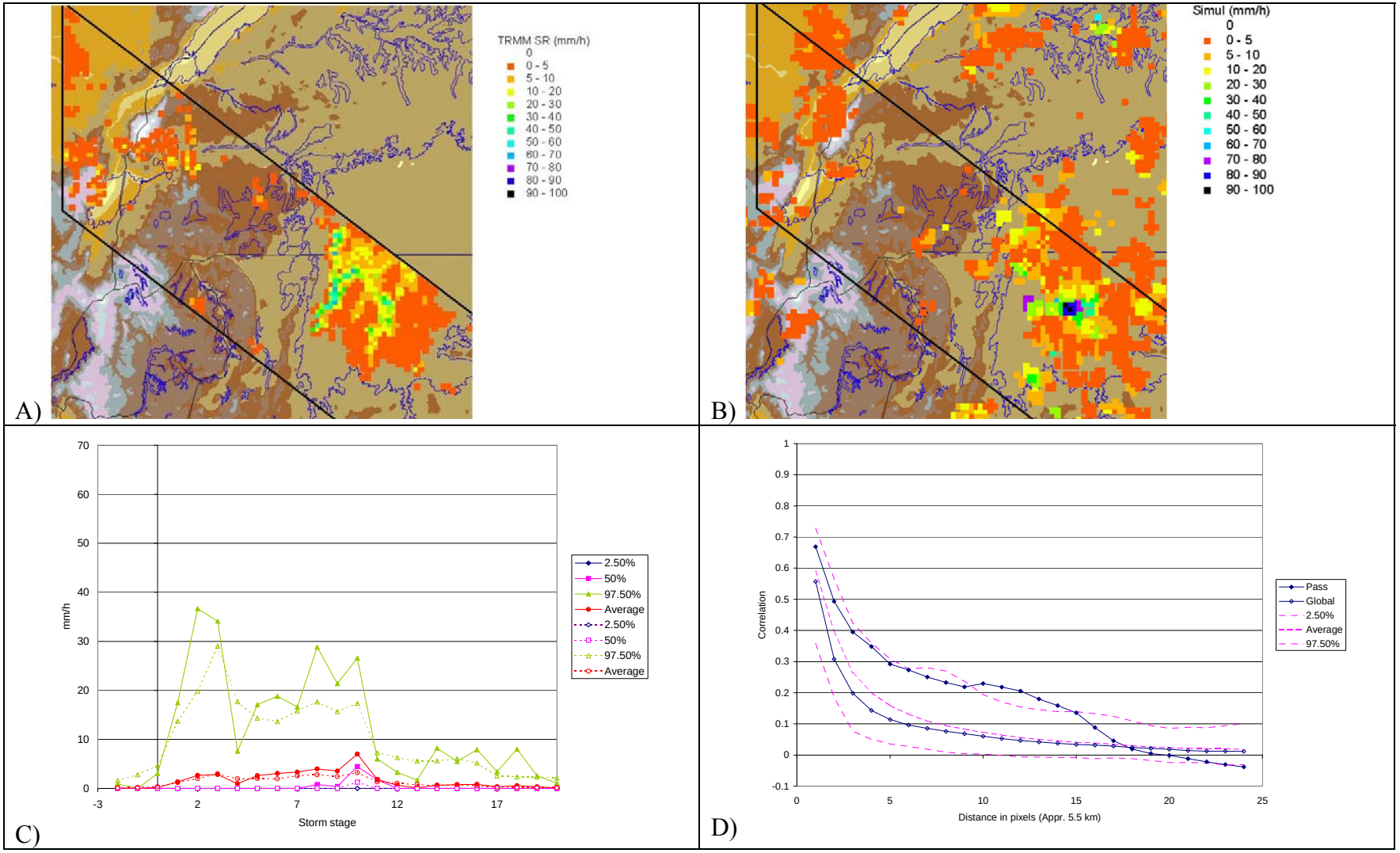


Figure 8.37 Simulation of single TRMM passes. A) Observed TRMM SR map; B) One realization of the TRMM rain map; C) Observed and model precipitation distribution. D) Observed and model spatial correlation. May 3, 1998 23:26 GMT.

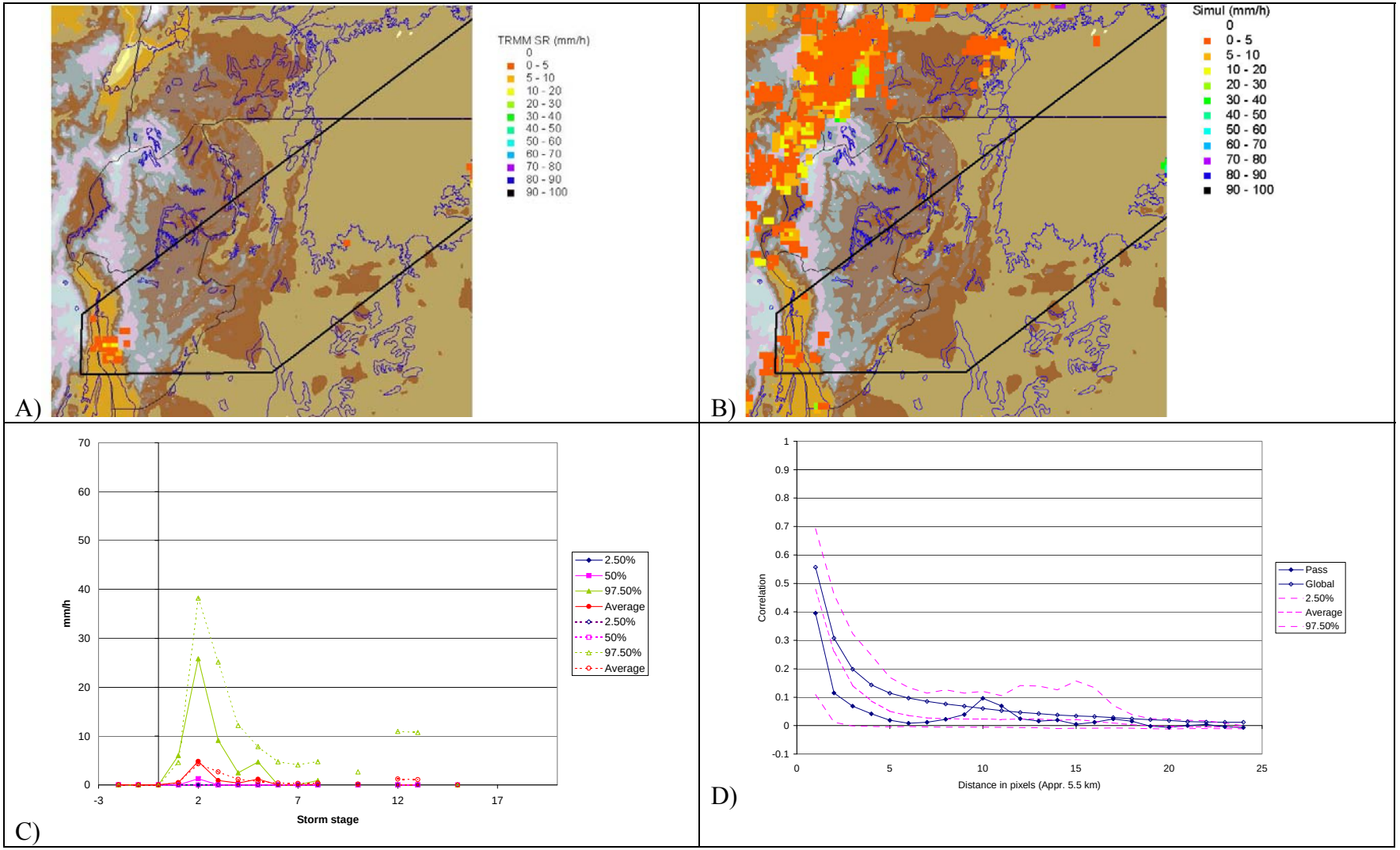


Figure 8.38 Simulation of single TRMM passes. A) Observed TRMM SR map; B) One realization of the TRMM rain map; C) Observed and model precipitation distribution. D) Observed and model spatial correlation. May 04, 1998 10:50 GMT.

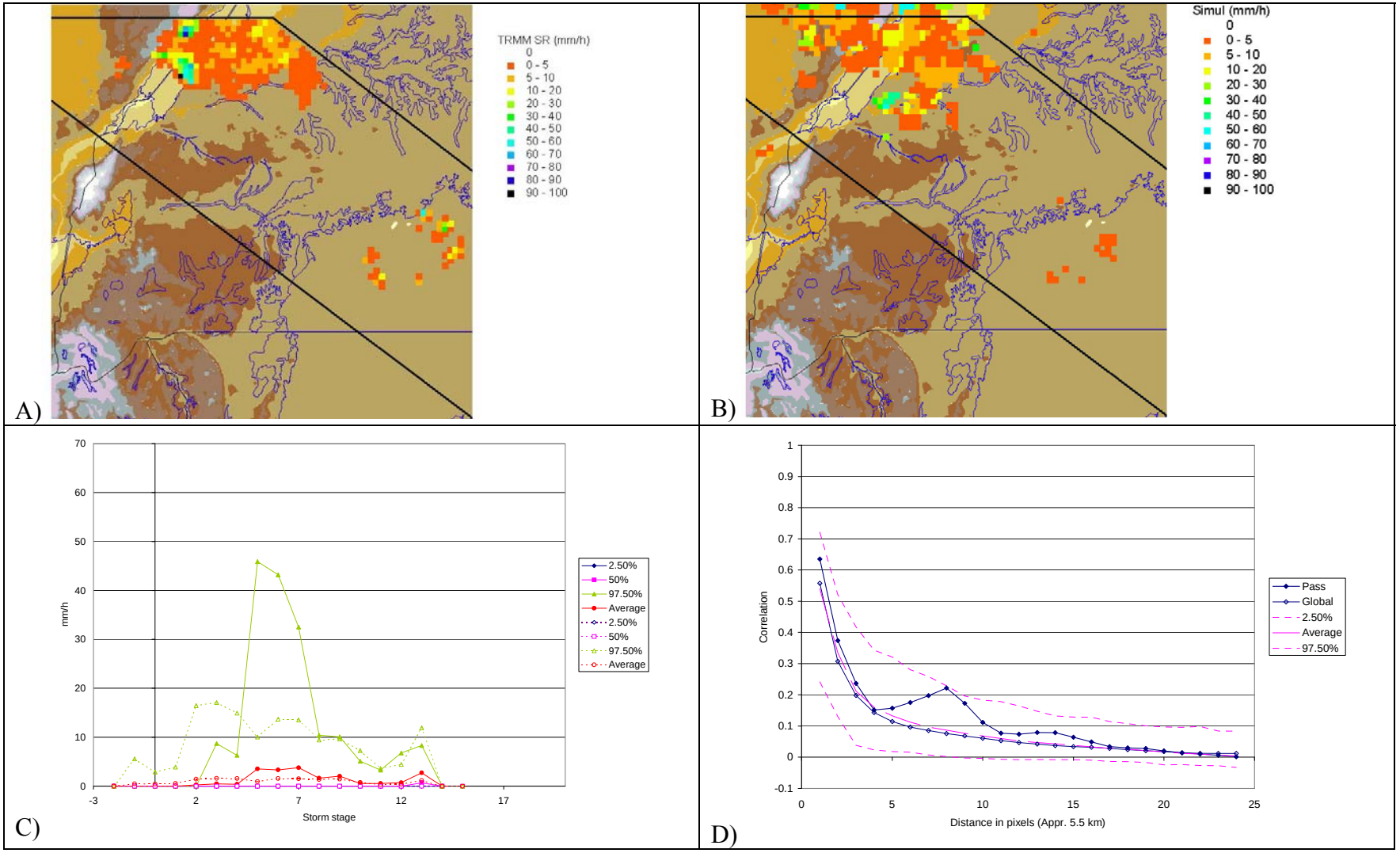


Figure 8.39 Simulation of single TRMM passes. A) Observed TRMM SR map; B) One realization of the TRMM rain map; C) Observed and model precipitation distribution. D) Observed and model spatial correlation. July 17, 1998 08:40 GMT.

8.6.2.3 Performances of the Temporally Uncorrelated Model

The temporally uncorrelated analytical-distribution model has been applied with the indicator residual kriging with 9-pixel F-extension and bias correction to estimate precipitation over the EKJ area according to the test format described at the beginning of the Chapter. The number of conditioning pixels has been increased to 15. Furthermore, the low-precipitation filtering has been increased from 0.5 to 0.7 mm h^{-1} as described in the previous chapter.

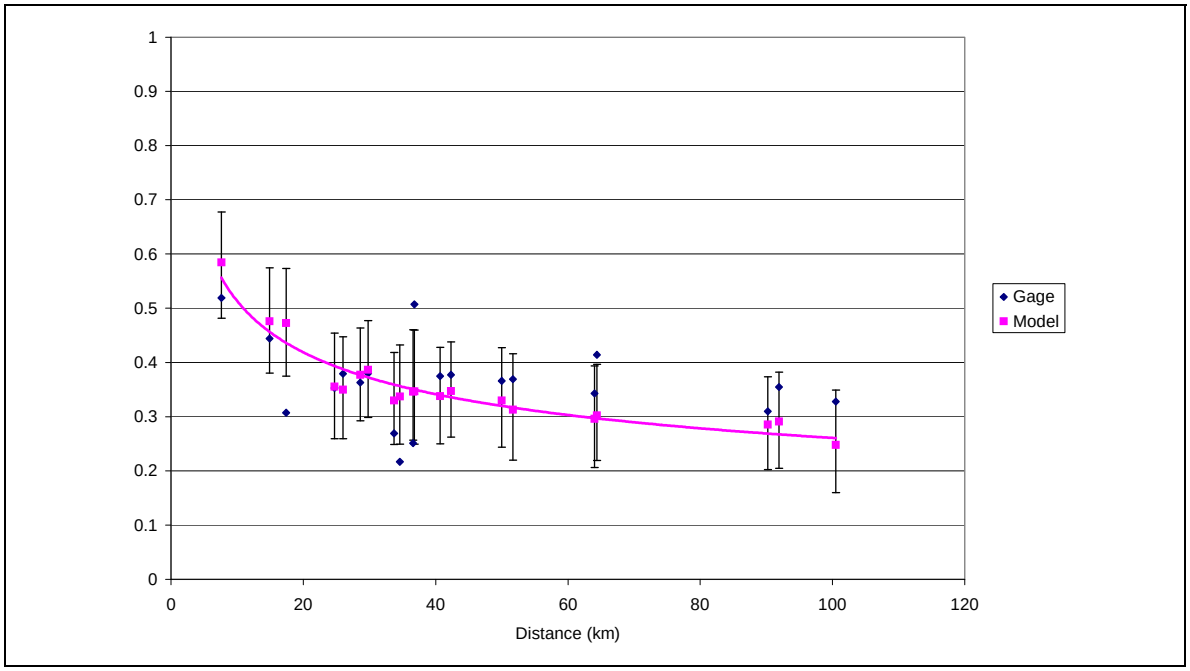


Figure 8.40 Correlation in daily precipitation for the EKJ area in 1996-1997. Temporally-uncorrelated analytical distributions with mixed kriging, 9-pixel F-extension, and bias correction.

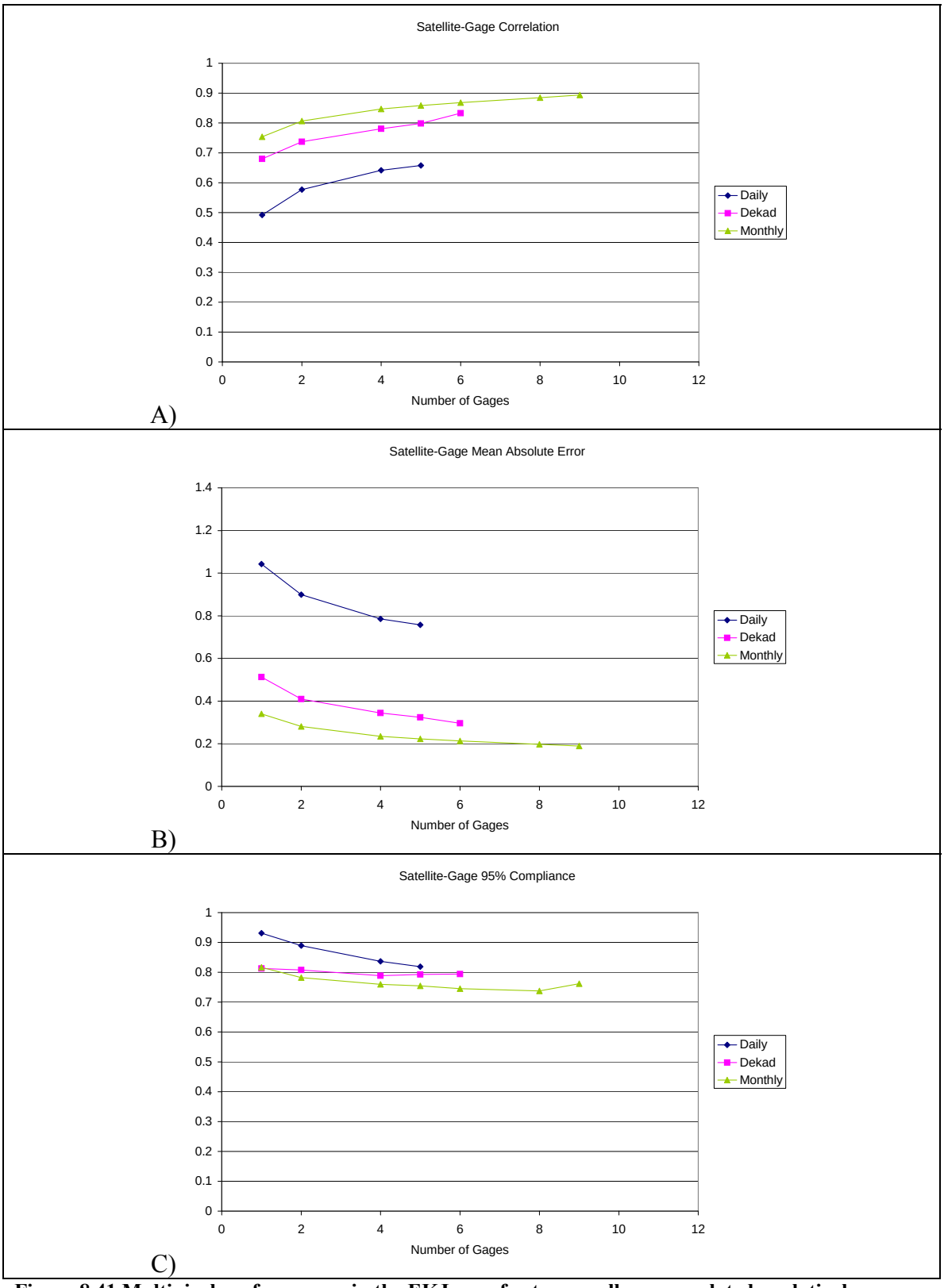


Figure 8.41 Multipixel performances in the EKJ area for temporally-uncorrelated analytical distributions with mixed kriging, 9-pixel F-extension, and bias correction.

The temporally uncorrelated analytical model has a spatial correlation slightly lower than its empirical counterpart, but it is still adequate to match the gage correlation (Figure 8.24 and Figure 8.40). This does not affect the 95% compliance rate, which, on the contrary, is better than the one featured by the empirical distributions (Figure 8.41). The lower bias of the analytical model is a possible cause of this contradictory result.

8.6.2.4 Performances of the Temporally Correlated Model

The temporally correlated analytical-distribution model has been applied with the indicator residual kriging, 9-pixel F-extension, and bias correction to estimate precipitation over the EKJ area according to the test format described at the beginning of the Chapter. The number of conditioning pixels has been increased to 15. It is reminded that in the analytical-distribution model, precipitation probability, conditional average precipitation rate, and conditional precipitation variance depend on the same pixel precipitation at *t-1* according to the model shown in Section 7.4.2. Therefore, at time *t*, the set of prior cdf parameters for one realization is chosen based on the precipitation field at time *t-1*, the cdf parameters at *t-1*, and the IR, VIS/WV, stage, and month at time *t*. The kriging procedure is then applied using this set of prior parameters as outlined in section 8.6.2.2. In addition, as in the uncorrelated precipitation simulation, rain below 1.0 mm h^{-1} has been discarded, considered as numerical noise. It is noted that, unlike the uncorrelated case, the low-rain filter has not been increased, because this operation would eliminate large areas of legitimate stratiform rain, which is often on the order of 1.0-1.5 mm h^{-1}.

Model performances are similar to those produced by the empirical distributions, rendering the comments expressed in Section 8.6.1.4 applicable in this case too. Therefore, this section will not repeat the same detailed analysis seen there, but will instead focus on the differences between the two models.

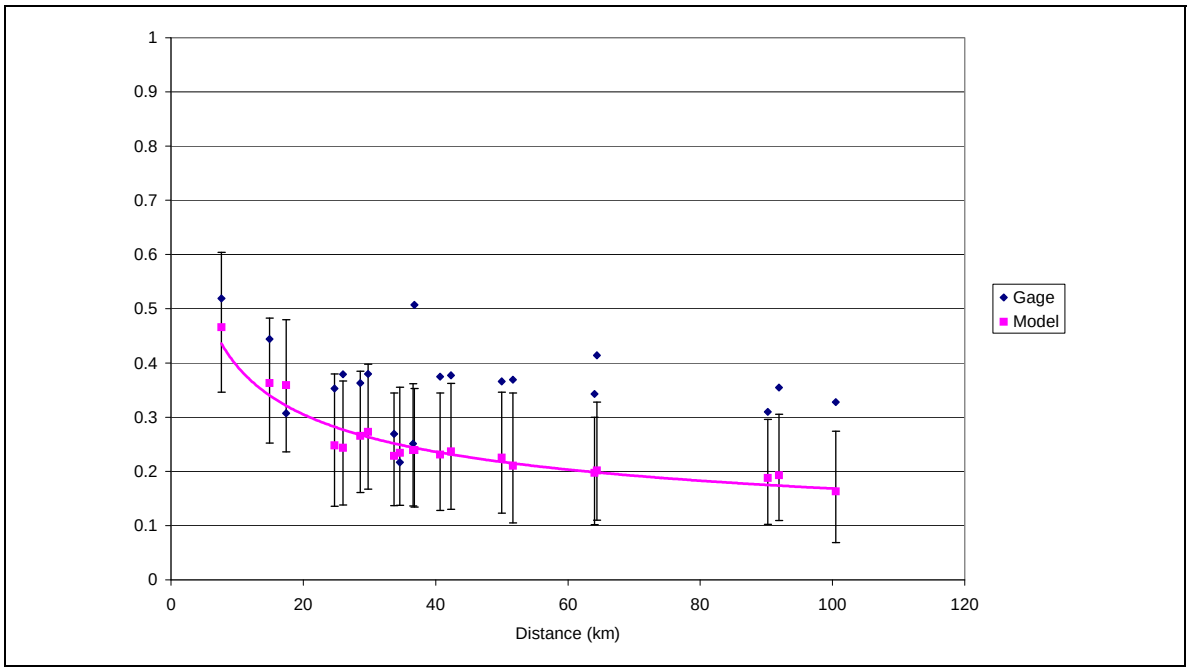


Figure 8.42 Correlation in daily precipitation for the EKJ area in 1996-1997. Temporally-correlated analytical distributions with mixed kriging, 9-pixel F-extension, and bias correction.

The analytical distributions have a spatial correlation that is 3-5 percentage points lower than that featured by the empirical distribution for most distances, and 7-8 percentage points lower for very short distances (Figure 8.26 and Figure 8.42). The spatially uncorrelated and the temporally-uncorrelated models based on the analytical distributions also feature lower spatial correlation than the equivalent models based on the empirical distributions (Sections 8.4, 8.6.1.3, and 8.6.2.3), but not to the same extent. The lower spatial correlation is here caused mainly by the bias-reduction procedure. In the absence of an increase of the low-rain filtering threshold during the simulation, this procedure has a stronger impact on the selected pixels.

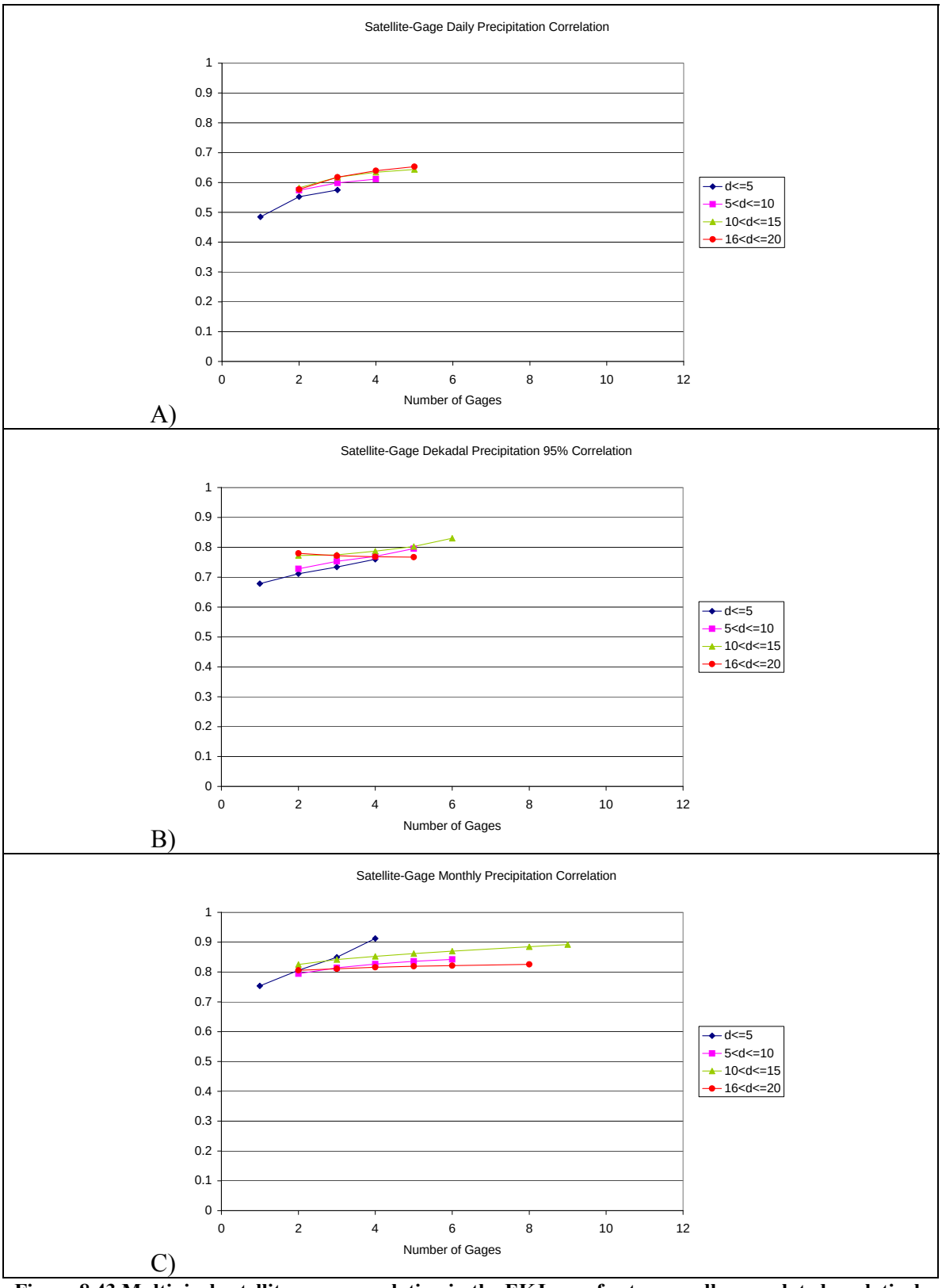


Figure 8.43 Multipixel satellite-gage correlation in the EKJ area for temporally-correlated analytical distributions with mixed kriging, 9-pixel F-extension, and bias correction.

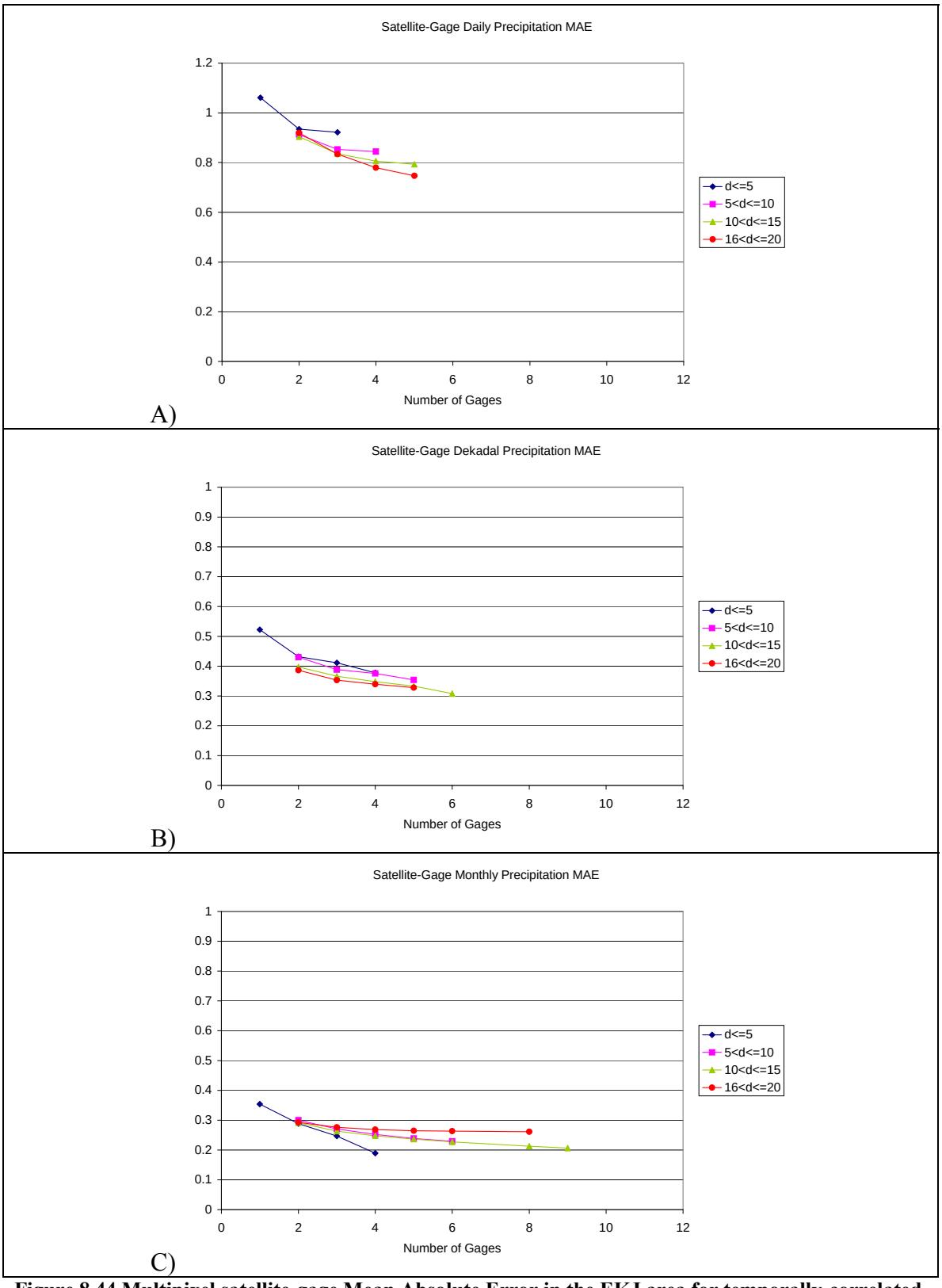


Figure 8.44 Multipixel satellite-gage Mean Absolute Error in the EKJ area for temporally-correlated analytical distributions with mixed kriging, 9-pixel F-extension, and bias correction.

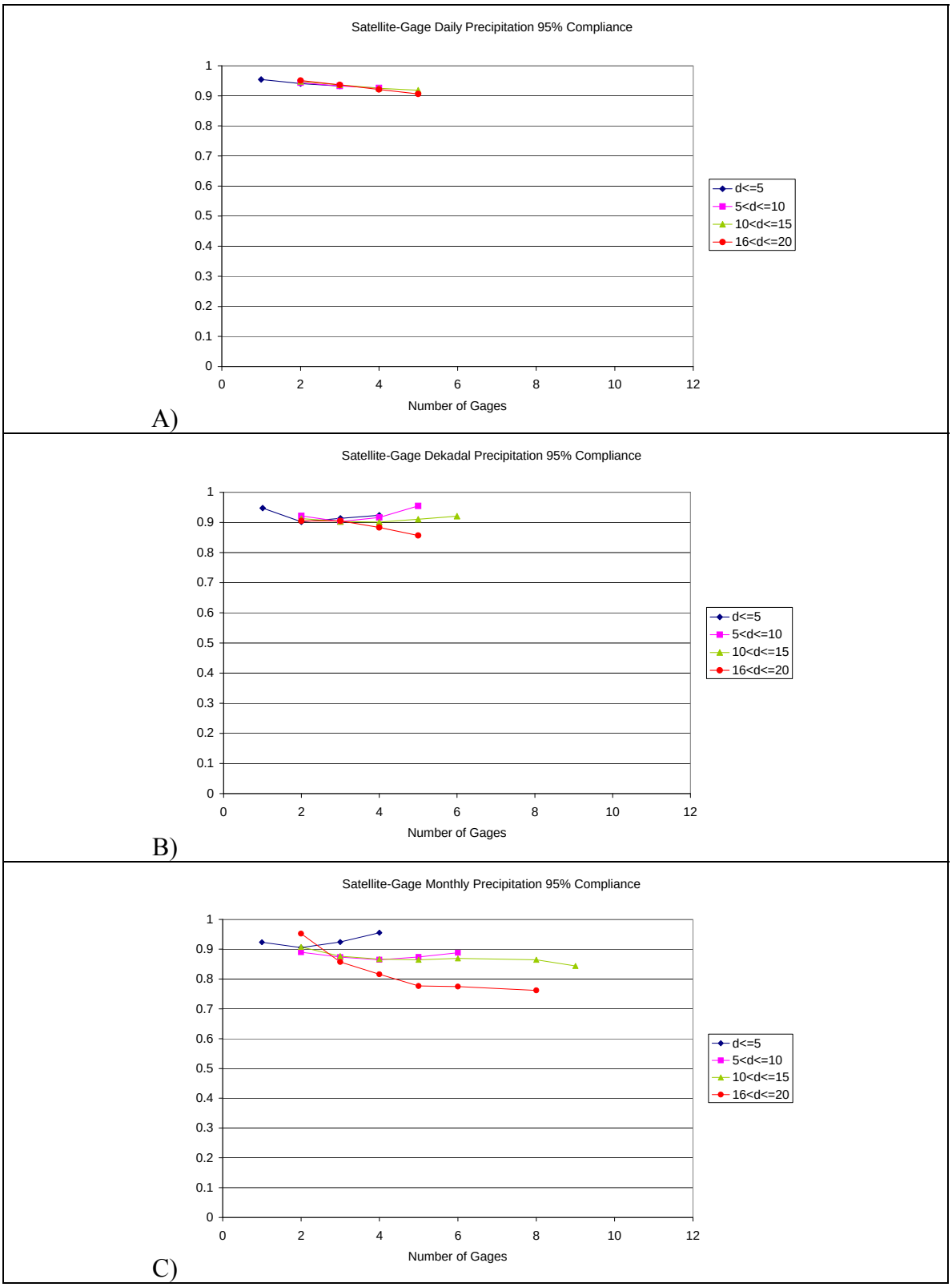


Figure 8.45 Multipixel satellite-gage 95% Compliance rate in the EKJ area for temporally-correlated analytical distributions with mixed kriging, 9-pixel F-extension, and bias correction.

Overall, correlation is good for distances below 20 km and acceptable for distances below 40 km. At higher distances, the low spatial correlation of the instantaneous precipitation is not able to smooth out the randomness introduced by the same pixel temporal correlation.

The analytical distribution model features a higher satellite-gage correlation than the empirical model because the distribution parameters depend also on the VIS/WV, and month (Figure 8.27 and Figure 8.43). The correlation is higher especially at the monthly resolution. Correspondingly, the MAE is lower at the monthly and decadal levels, but it is slightly higher at the daily level, probably as a side effect of the bias-reduction procedure. Compliance rates are slightly higher than those of the empirical function model. The higher compliance rates of the analytical distributions compensate for the lower spatial correlation.

Table 8.6 reports the reliability of the analytical distribution model, according to the same methodology of Section 8.6.1.4. The reliability of the two models is essentially identical.

Table 8.6 95% Compliance of the temporally correlated, spatially correlated analytical distribution model.

Size of square	Daily	Dekadal	Monthly
<=0.25°	0.93	0.90	0.90
<=0.50°	0.93	0.91	0.87
<=0.75°	0.92	0.90	0.84
<=1.00°	0.91	0.90	0.84

Estimation uncertainty is also very similar to the empirical model (Figure 8.46), with a slightly larger reduction in the estimation uncertainty at the monthly resolution.

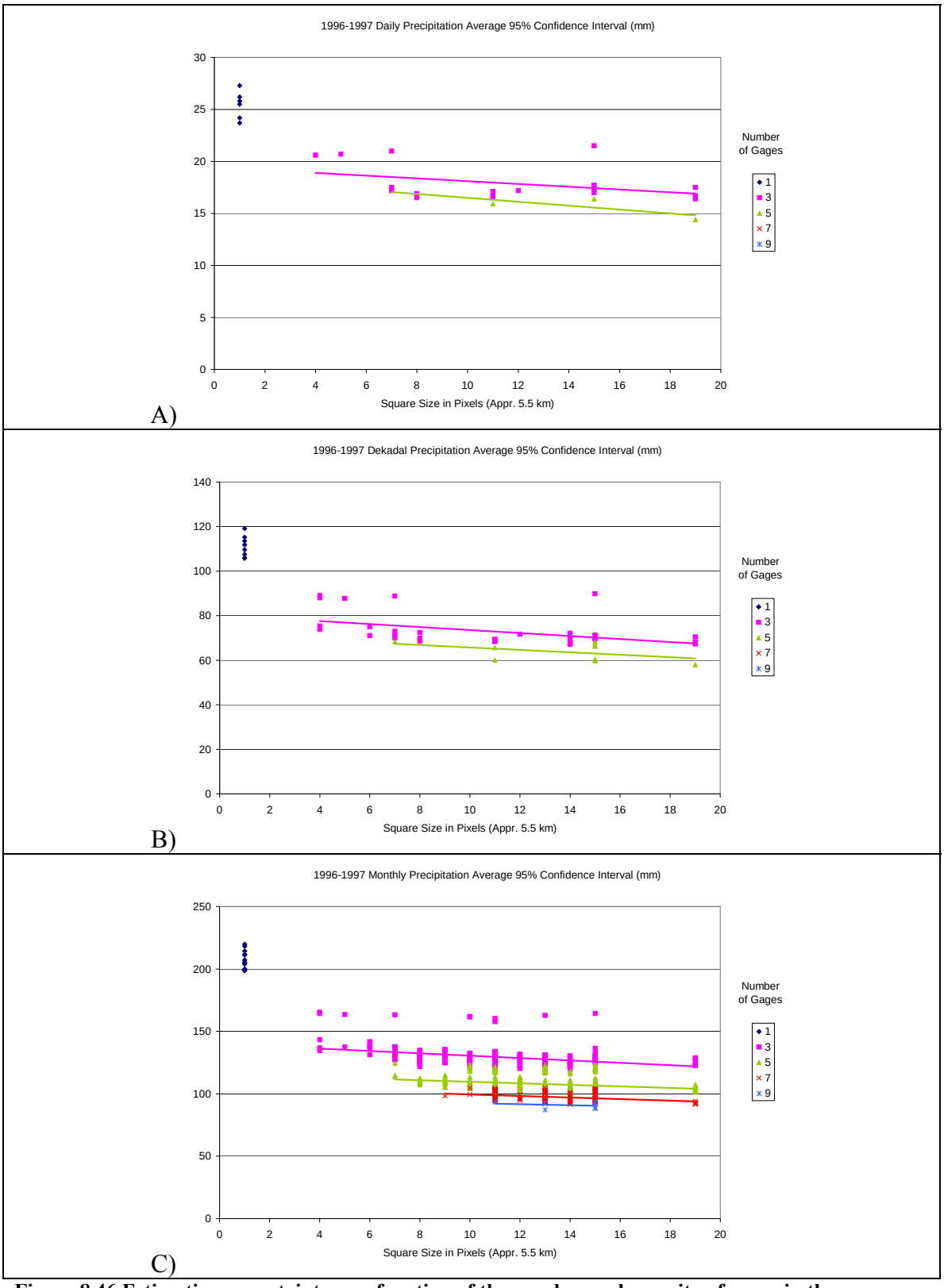


Figure 8.46 Estimation uncertainty as a function of the number and sparsity of gages in the combination. Temporally correlated analytical distributions with mixed kriging, 9-pixel F-extension, and bias correction

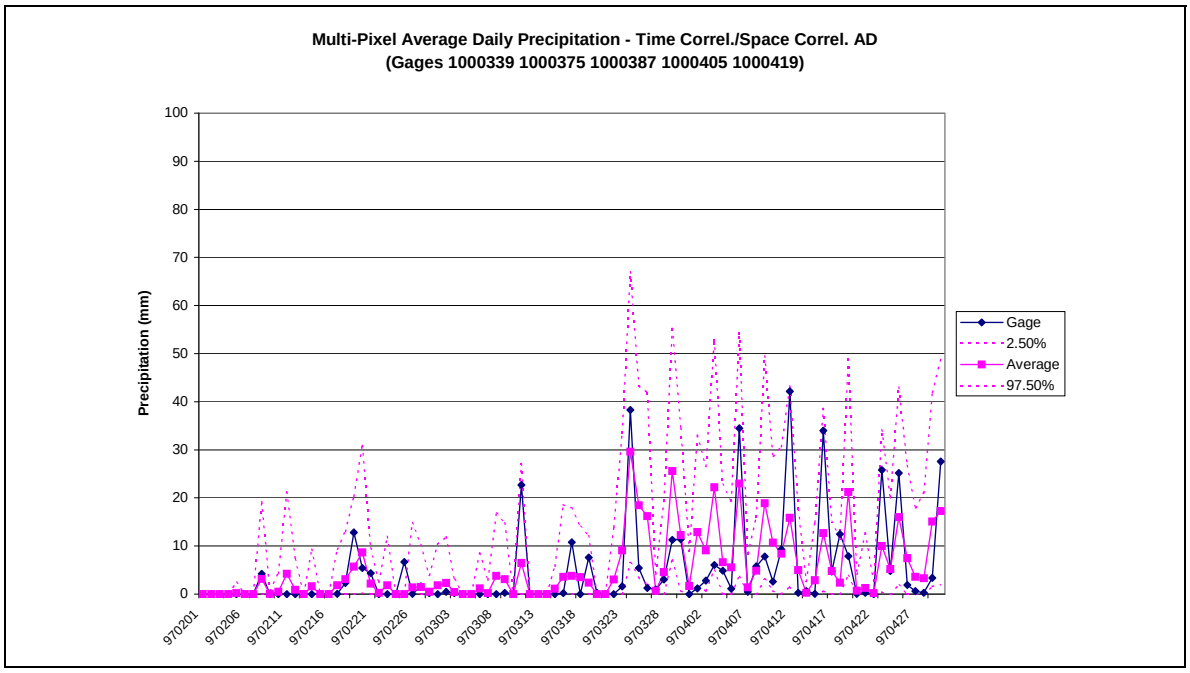


Figure 8.47 Daily precipitation at a combination of five gages in the EKJ area. Temporally correlated analytical distributions Mixed kriging, 9-pixel F-extension, and bias correction.

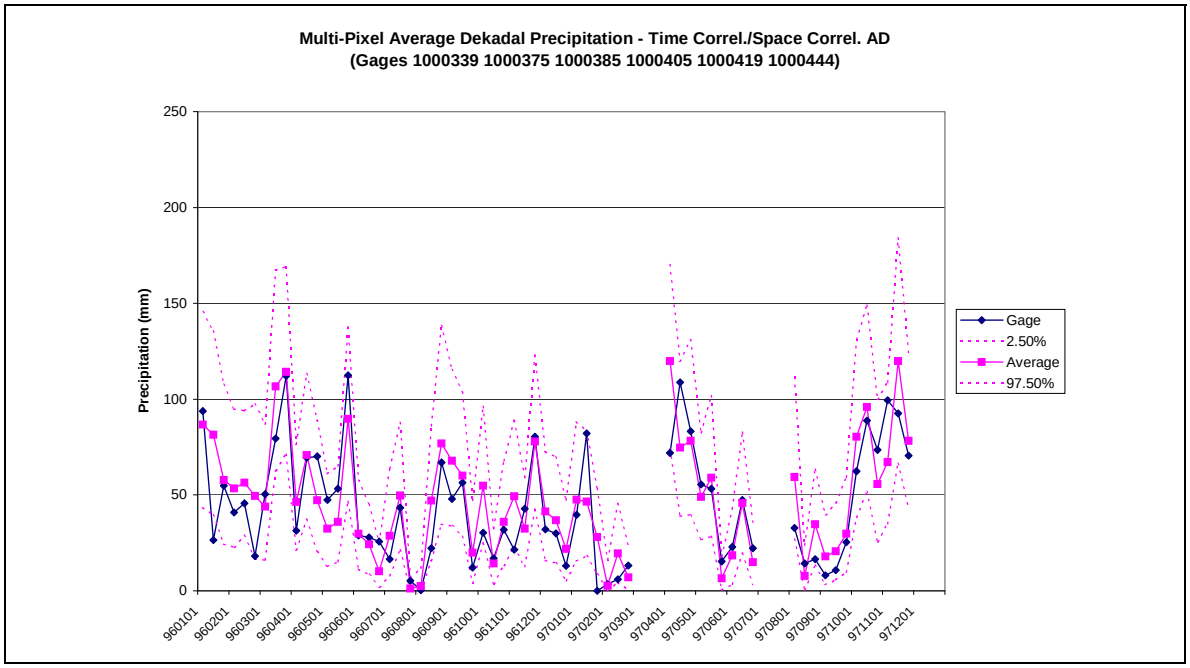


Figure 8.48 Dekadal precipitation at a combination of six gages in the EKJ area. Temporally correlated analytical distributions. Mixed kriging, 9-pixel F-extension, and bias correction.

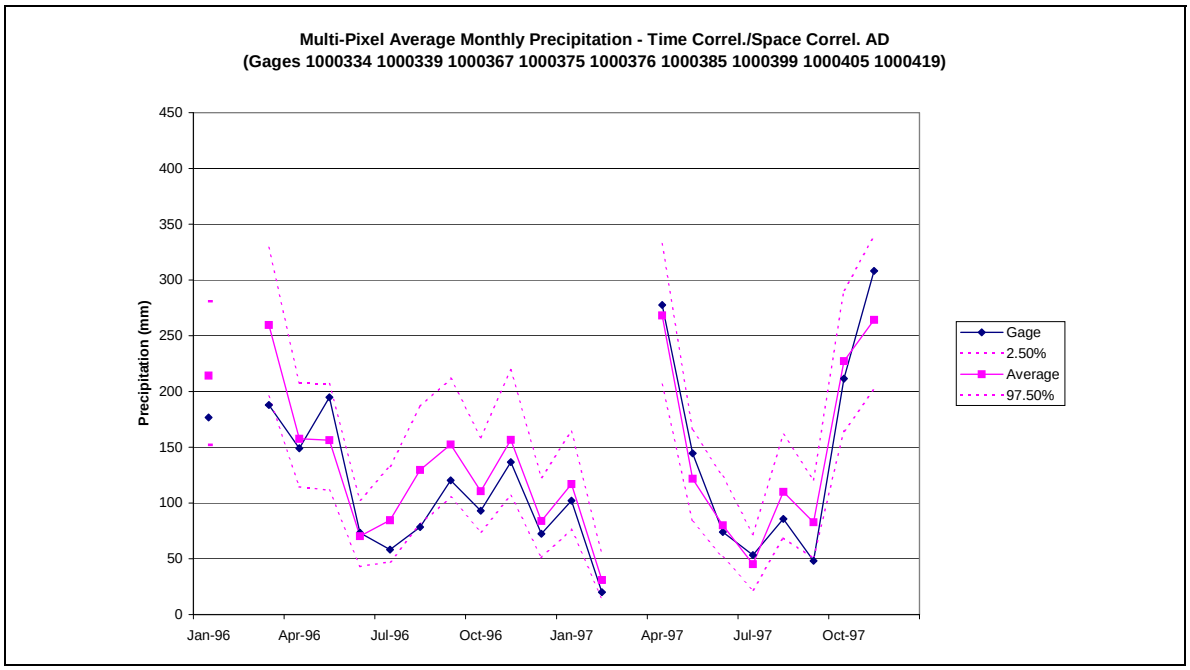


Figure 8.49 Monthly precipitation at a combination of nine gages in the EKJ area. Temporally correlated analytical distributions. Mixed kriging, 9-pixel F-extension, and bias correction.

Model results (Figure 8.47, Figure 8.48, Figure 8.49) show higher skill of the analytical model to estimate monthly and decadal precipitation, while at the daily level the skill is similar. It is noted, however, that the confidence intervals produced by the analytical functions for this combination of gages is smaller.

Figure 8.50 shows the variation in daily mean areal precipitation and its uncertainty as a function of the size of the area considered in the MAP for the two instances reported in Figure 8.51. Starting at the pixel at the center of the square, the area used for computing the MAP was gradually increased until it reached the size of the square of interest (1.0ºx1.0º). Computing the MAP with or without spatial correlation yields average estimated MAP, (AIVWNCC) and (AIVWNCU) respectively, that are very similar to each other. They both show some relevant initial variation when the area used for computing the precipitation is small, but quickly stabilize over larger areas.

Conversely, the estimation uncertainty, represented in Figure 8.50 by the ratio between the width of the 95% confidence interval and the MAP, decreases much more

rapidly when the spatial correlation is not considered. Results reported in the earlier part of this chapter showed that this rapid decline results in an underestimation of the precipitation uncertainty and, consequently, in unreliable confidence intervals. The incorporation of the precipitation spatial correlation, on the other hand, maintains higher estimated precipitation uncertainty reflecting the true estimation error. Notably, the true uncertainty of the daily MAP is above 100% even for large areas.

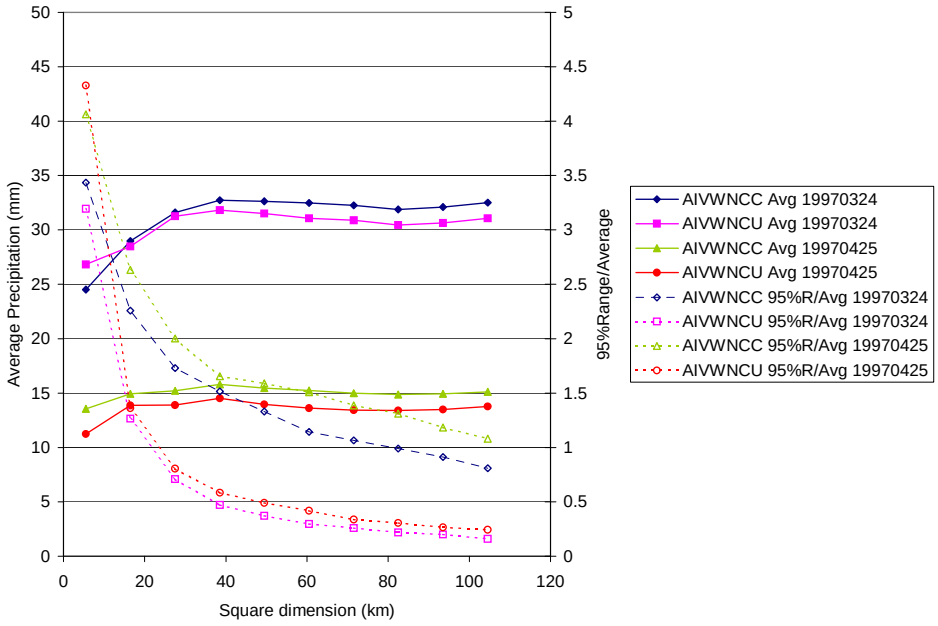


Figure 8.50 Estimated MAP and ratio 95%-Confidence interval width/MAP as a function of the area size. Spatially correlated (AIVWNCC) and spatially uncorrelated (AIVWNCU) models.

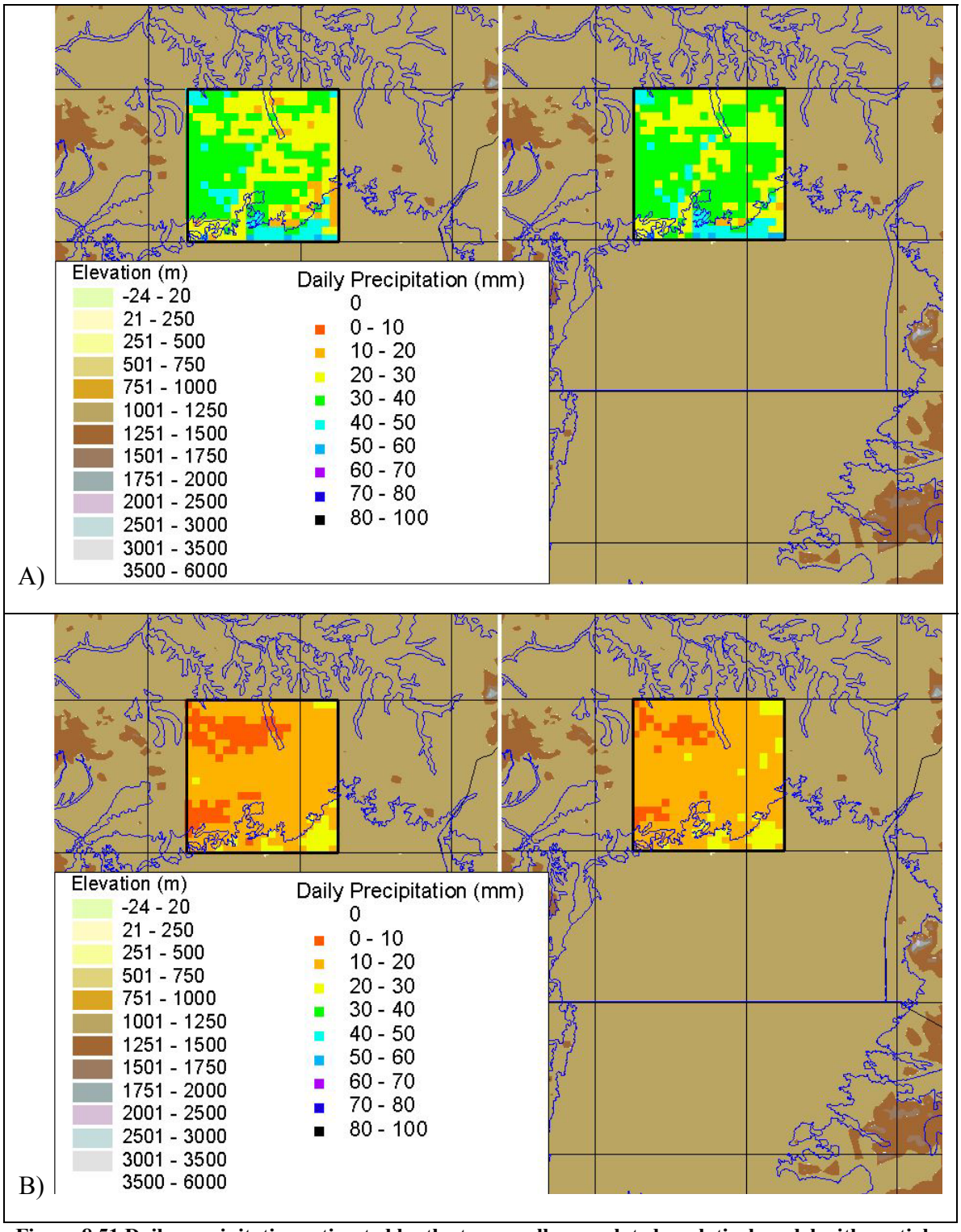


Figure 8.51 Daily precipitation estimated by the temporally correlated analytical model with spatial correlation (Right) or without (Left). A) March 23, 1997. B) April 25, 1997.

8.7 Summary

The aggregation of simulated precipitation over areas containing many pixels produces mean areal precipitation estimates that, on average, agree with gage measurements. Correlation increases and error decreases with the number of pixels considered. On the other hand, the simple aggregation of single-pixel precipitation ignores the rainfall spatial correlation, underestimating precipitation variability. Two procedures based on sequential simulation have been implemented for generating spatially correlated precipitation fields to correct this effect. For the empirical distribution model, the spatial correlation is introduced using indicator kriging with spatially varying prior means. The procedure for the analytical distributions is instead a two-step algorithm: the first step simulates the rain/no-rain distribution using indicator kriging; the second step simulates the lognormal distribution of precipitation intensity using non-stationary simple kriging. In temporally-uncorrelated settings, the procedures are successful in reproducing the average spatial correlation of TRMM samples and the spatial correlation of gage-recorded precipitation. The single-pixel temporal correlation models presented in Chapter 7 cannot replicate the spatial correlation of the simulated precipitation field, because of the strong pixel-by-pixel dependencies. The two kriging algorithms satisfactorily compensate for this discrepancy only for distances below 40 km. At longer distances, the instantaneous precipitation spatial correlation is not sufficiently strong to ameliorate this problem. In spite of this, the kriging procedures greatly increase the capability of the models to represent the precipitation variability at all temporal and spatial resolutions, producing reliable estimates for most scales of interest.

The simulation of the precipitation rain field could be improved by decreasing the model spatial resolution from single 0.05°x0.05° pixels to 0.15°x0.15° aggregates. This operation would yield several positive outcomes:

- The spatial distribution of the a-priori parameters would be smoother;

- The precipitation spatial correlation would be stronger;
- The temporal correlation would have a smaller impact;
- The a-priori precipitation distribution would be closer to normal with lower probability of no-rain.

Other performance improvements could be achieved by separating the simulation of large storms, which are characterized by stronger spatial correlation functions, from that of the smaller storms, which are less strongly correlated.

CHAPTER 9

CONCLUSIONS AND RESEARCH RECOMMENDATIONS

9.1 Conclusions

This research makes three important contributions in the area of precipitation estimation using satellite and conventional rain gage data.

First, a novel procedure was developed to identify the presence and temporal evolution of convective storms by detecting their IR signal. This is accomplished by examining a three-hour IR window throughout the day and recognizing the forms typical of the onset of convective storms. The form recognition is accomplished by a neural network trained on historical IR data associated with convective events as identified by the TRMM precipitation radar (TRMM-PR). In the area of application (Lake Victoria, East Africa), this convective pattern recognition procedure exhibits a probability of detection of 0.8 and a false alarm rate of 0.05.

Furthermore, the temporal distance from the neural network's detection of the convective storm's onset was used to track the temporal evolution of the detected convective patterns (storm stages). The storm stages are shown to improve the precipitation estimation at a single pixel but also provide the basis for deriving the precipitation temporal correlation from the spatial distribution of precipitation reported in TRMM images. Storm stage classification also enables the implementation of an ergodic working principle according to which TRMM spatial images provide information on the temporal storm evolution.

Second, detailed relations between precipitation rates and storm stage, IR, VIS/WV, terrain, and season data were constructed using a multi-year database of coincident TRMM-PR and geostationary images. Such relations are shown to

characterize with sufficient accuracy the average precipitation regimes and proved to be useful in environments where orography changes markedly within few dozens of kilometers (such as the Lake Victoria Basin). In comparisons against a large database of precipitation records including more than one hundred rain gages distributed over regions with different climates and spanning three years, the satellite-gage bias of the precipitation estimation developed herein is of the order of 5-10%. This figures are much less than those of the Global Precipitation Index, GPI, and the 3B42, modified GPI index (Adler et al., 2000), which respectively amount to ~70-75% and ~ 50%. Satellite to gage correlation is also higher, especially at the daily resolution (3-5 percentage points). Precipitation estimation in areas with high relief or areas of transition between land and open water are less reliable (often overestimated), but less than the equivalent measurement by GPI.

Figure 9.1 shows that the procedure developed herein matches very well the 1996-1997 cumulative precipitation estimate obtained independently by a method for interpolating rain gages in (Mitchell et al., 2005). The most notable discrepancies are located over the Congo plateau and over Lake Victoria. In the case of the Congo plateau, the differences may be attributed to the high cloudiness of that part of the Congo plateau but also to the questionable quality and quantity of rain gage data due to civil unrest. In the case of Lake Victoria, however, Mitchell et al. (2005) obtained the precipitation by interpolating data from gages along the lake shoreline, leading to a clear underestimation of precipitation. In this respect, the satellite-based estimate of precipitation over the lake matches well the average annual precipitation trends reported by other authors (Shahin, 1985) and is more reliable. Figure 9.1 exemplifies the importance of satellite precipitation estimation for water resources management in this part of the world. Increasing the size of the database of coincident TRMM-Geostationary database from the current two years to all available years promises to yield even better results, allowing for more spatial and temporal differentiation of the precipitation relations.

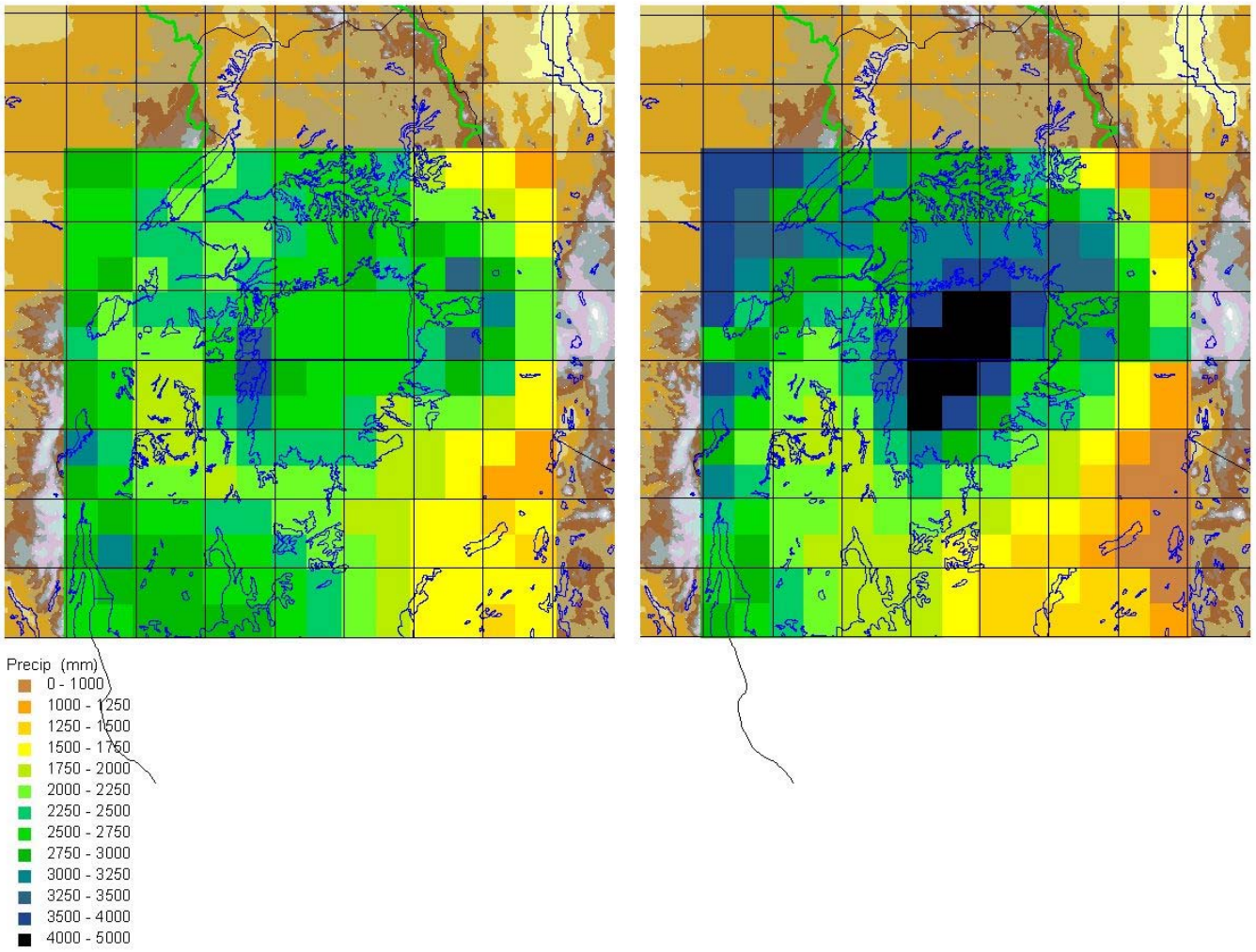


Figure 9.1 Estimated cumulative precipitation for 1996-1997. Left, interpolated gage records (After Mitchell et al., 2005). Right, AIWVNUU estimation.

Third, and most important, the methodology developed in this research quantifies the uncertainty associated with precipitation estimates. Most traditional procedures for estimating precipitation do not provide information on the uncertainty affecting their precipitation estimates, despite its demonstrated presence and magnitude (Section 2.3.3.4). A few existing procedures supply such information but normally report just the mean and variance of precipitation at each elementary spatial and temporal element (pixel and time step). However, such uncertainty characterization is not adequate for three reasons:

1. Precipitation at fine temporal and spatial resolution has markedly skewed distributions, often of a mixed discrete/continuous nature, reducing the significance of mean and variance.
2. While averaging the elemental mean precipitations of a number of pixels and time-steps determines the average precipitation over larger areas and/or periods, evaluating the associated uncertainty from the elemental variances is impossible because of the precipitation spatial and temporal correlation.
3. The precipitation estimates are commonly used as input to some hydrologic model, the output of which is a non-analytical function of the input. The evaluation of the output uncertainty is thus possible only in a Monte Carlo context.

The approach pursued in this research was to provide an ensemble of equally probable precipitation estimates for each elemental unit. The elements of the ensemble were obtained by sampling the full distribution of TRMM PR data associated with the combination of the IR, VIS/WV, storm stage, month, and terrain characterizing the elemental unit. Furthermore, the generation of the single precipitation realizations was constrained to observe some model of the precipitation spatial and temporal correlation.

Accordingly, the ensemble of precipitation estimates obtained by aggregating elemental precipitation values from the same realization properly characterizes the estimate uncertainty over any spatial and temporal scale of interest, as shown in Chapter 8.

Section 6 showed that the distributions of half-hour precipitation rates are skewed and cannot be represented by standard distributions. Two approaches were followed to address this problem: 1) precipitation distributions were represented using sample (or empirical) distributions, and 2) precipitation distributions were modeled by a discrete probability distribution for no-rain and a continuous lognormal probability distribution for rain intensity assuming that a "rain" event occurred.

The first approach does not involve any assumption on the precipitation distribution and proved to be more robust, but considerably cumbersome. Specifically, this approach requires more computer memory and execution time, especially in cases where spatial and temporal correlations need be considered.

The analytical approach is less robust with the limited database of precipitation rates available in this study, but faster and easier to implement, even in cases where spatial and temporal correlation needs to be incorporated. Furthermore, the analytical approach is more amenable to the assimilation of new satellite and gage data and to the localization in space and/or time of the relation between precipitation and satellite information.

Chapter 7 showed that the representation of estimate uncertainty over periods longer than one time-step is acceptable only when the estimation takes into consideration the precipitation temporal correlation. The precipitation temporal correlation was introduced herein via a single-pixel one-lag Markov process. In this way, the procedure is able to account for most of the precipitation temporal variability, yet at the expense of estimation spatial correlation.

Likewise, Chapter 8 showed that the representation of estimate uncertainty over areas larger than one pixel is possible only when the estimation takes into consideration

the precipitation spatial correlation. In this research the precipitation spatial correlation was implemented using a sequential simulation approach that includes kriging.

This procedure is able to reproduce well the average spatial correlation of temporally uncorrelated precipitation. Conversely, it is unable to reliably reproduce the spatial correlation of the more organized larger storms. Further, kriging is not able to smooth out the disruptions introduced by the temporal correlation at medium-large distances. Nevertheless, the procedure greatly increases the reliability of the uncertainty characterization, especially for areas smaller than 0.5ºx0.5º (~3000 km^2).

One of the goals set forth in this research was to create a probabilistic model of precipitation using data readily available in the developing world, such as TRMM and geostationary data and daily precipitation records from sparse networks. The results show that it is possible to accomplish this goal even with limited resources and data sparse environments.

9.2 Future Research Recommendations

The methodology developed in this research can be improved and expanded in several areas. Expanding the database of collocated TRMM and geostationary images would greatly improve the accuracy of precipitation estimates by allowing a better characterization of the relations between precipitation rates and IR, VIS/WV, storm stage, terrain, and month. Further, a larger database could allow discarding the TRMM and geostationary images that are not closely coincident (i.e., that are between ten and fifteen minutes apart). In this respect, the use of the less precise, but more frequently available, precipitation estimates from passive microwave sensors aboard TRMM and DMSP satellites should be evaluated. In addition, the higher temporal frequency of the passive microwave sensor makes them better candidates than the TRMM PR data for their direct assimilation into a precipitation estimation scheme. In the future, the deployment of the

new generation of TRMM satellites (GPM) is also expected to provide much better spatial and temporal precipitation coverage and enhance the results obtained herein.

Second, under the availability of more data, shallow convection storms could separately be detected using a second neural network trained to identify such events at mid and high elevations (Section 2.5).

Third, a better characterization of the precipitation temporal correlation should be pursued by rendering the precipitation at time t depended upon previous precipitation values over areas larger than a single pixel. Unfortunately, TRMM images do not provide enough information for the development of such extended relations. Such high temporal and spatial relations require radar or sufficient rain-gage data.

Fourth, two measures are recommended to further improve the consideration of spatial correlation: First, the model spatial resolution should be reduced from 0.05ºx0.05º to 0.15ºx0.15º. The resulting resolution is still less than or equal to that adopted by most estimation procedures (e.g., Bellerby, 2000; Huffman et al., 2001), but it should help to (a) improve the spatial and temporal correlation of precipitation, (b) reduce the variability of the unconditional average precipitation, and (c) increase the "normality" of the precipitation distribution.

A second measure that would improve the precipitation spatial characterization is to differentiate the spatial correlation of small storms from that of larger storms. These two types of events could be separated by using some readily available indexes of cloud dynamics over larger regions, such as the average IR or the fraction of pixels with IR below a certain threshold.

Fifth, direct assimilation of measured precipitation, both from gage and from satellite, should also be studied as well the localization in space and time of the precipitation distributions by comparing the unconditional and measured precipitation

mean and variance for similar conditions and correcting for the differences using, for example, a Bayesian scheme.

Lastly, smoothing the distribution of the model parameters or using other more complex algorithms to compute the probability of rain, and the mean and variance of the conditional rain rate could improve the estimation and allow for a better synergistic usage of coincident microwave and geostationary data.

REFERENCES

Adler, R.F., Huffman, G.J., Bolvin, D.T., Curtis, S., and Nelkin, E.J (2000). Tropical Rainfall Distributions Determined Using TRMM Combined with Other Satellite and Rain Gage Information. J. Appl. Meteor. 39: 2007 – 2023.

Adler, R.F., Kidd, C., Petty, G., Morrisey, M., and Goodman, M. H., (2001). Intercomparison of Global Precipitation Products: The Third Precipitation Intercomparison Project (PIP-3). Bull. Amer. Meteor. Soc., 82, 1377–1396.

Adler. R.F., and Mack, R.A.. (1984). Thunderstorm Cloud Height-Rainfall Rate Relations for Use with Satellite Rainfall Estimation Techniques. J.Climate Appl. Meteor. 23: 280-296.

Adler. R.F., and Negri, A.J. (1988). A Satellite Infrared Technique to Estimate Tropical Convective and Stratiform Rainfall. J.Climate Appl. Meteor. 27: 30-51.

Adler, R.F., Negri, J.A., Keehn, P.R., and Hakkarinen, I.M. (1993). Estimation of Monthly Rainfall Over Japan and Surrounding Waters from a Combination of Low-orbit Microwave and Geosynchronous IR Data. J. Appl. Meteor. 32: 335-356.

Anagnostou, E.N. (2004). Overview of Overland Satellite Rainfall Estimation for Hydro-Meteorological Applications. Surveys in Geophysics 25: 511–537, 2004.

Arkin, P.A. (1979). The Relationship Between Fractional Coverage and Rainfall Accumulations During GATE over B-scale Array. Mon. Wea. Rev. 107: 1382-1387

Arkin, P.A. and Meisner, B.N. (1987). The Relationship Between Large-scale Convective Rainfall and Cold Cloud Over the Western Hemisphere During 1982-84. Mon. Wea. Rev. 115: 51-74

Arkin. P.A., and Xie, P. (1994). The Global Precipitation Climatology Project: First Algorithm Intercomparison Project. Bull. Amer. Meteor. Soc.: 75(3): 401-419.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a). Artificial neural networks in hydrology. I: preliminary concepts. Journal Of Hydrologic Engineering, 5(2), 115-123.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000b). Artificial neural networks in hydrology. II: hydrologic applications. Journal Of Hydrologic Engineering, 5(2), 124-137.

Atkinson, P.M., and Lloyd,7 C.D. (1998). Mapping Precipitation in Switzerland with Ordinary and Indicator Kriging. J. Geographic Information and Decision Analysis 2(2): 65-76.

Atlas, D., and Bell, T.L. (1992). The Relation of Radar to Cloud Area-Time Integrals and Implications for Rain Measurement from Space. Mon. Wea. Rev. 120(9): 1997-2008.

Austin, P.M. (1987). Relation Between Radar Reflectivity and Surface Rainfall. Monthly Weather Review 115: 1053 – 1070.

Barancourt, C., Creutin, J.D., and Rivoirard, J. (1992). A Method for Delineating and Estimating Rainfall Fields. Water Resour. Res. 28 (4): 1133-1144.

Bell, T.L. (1987). A Space-Time Stochastic Model of Rainfall for Satellite Remote-Sensing Studies. J. Geophys. Res. 92(D8), 9631-9643.

Bell, T.L., Abdullah, A., Martin, R.L., and North, G.R. (1990). Sampling Errors for Satellite-Derived Tropical Railfall: Monte Carlo Study Using a Space-Time Stochastic Model. J. Geophys. Res. 95(D3): 2195-2205.

Bellerby, T.J., and Barrett, E.C. (1993). A Strategy for the Calibration by Collateral Data of Satellite Estimates for Shorter Periods. J. Appl. Meteor. 32: 1365-1378.

Bellerby, T., Todd, M., Kniveton, D., and Kidd, C. (2000). Rainfall Estimation from a Combination of TRMM Precipitation Radar and GOES Multispectral Satellite Imagery Trough the Use of an Artificial Neural Network. J. Appl. Meteor. 39: 2115 – 2128.

Bras, R.L. (1990). Hydrology: An Introduction to Hydrologic Science. Addison-Wesley, Reading, MA.

Caers, J. (2000). Direct Sequential Simulation. Proceedings of the 6th International Geostatistics Congress. Cape Town, RSA..

Canada Centre for Remote Sensing (2002). Fundamentals of Remote Sensing Tutorial, http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/tutore.html (Accessed in May 2002).

Carlson. P.E. and Marshall, J.S. (1972). Measurement of Snowfall by Radar. J. Appl. Meteorol. 11: 494-500.

Churchill, D.D. and Houze, R.A. (1984). Development and Structure of Winter Monsoon Cloud Clusters on 10 December 1978. J. Atmos. Sci. 41(6): 933 -960.

D'Souza, G., Barrett, E.C., and Power, H. (1990) Satellite Rainfal Estimation Techniques Using Visible and Infrared Imagery. Remote Sensing Reviews 4(2): 379-414.

Deutsch, C.V., andd Journel, A.G. (1998). GSLIB: Geostatiscical Software Library and User's Guide. Oxford University Press, New York, NY.

Dingman, S.L. (2002). Physical Hydrology. Second Edition. Prentice Hall, Upper Saddle River, N.J.

Doneaud, A.A., Smith, P.L., Dennis, A.S., and Sengupta, S. (1981). A Simple Method for Estimating Precipitation Over an Area. Water Resources Research 17(6): 1676-1682.

Ebert, E.E., Manton, M.J., Arkin, P.A., Allam, R.J., Holpin. G.E., and Gruber, A. (1996). Results from the GPCP Algorithm Intercomparison Programme. Bul. Am. Meteor. Soc. 77(12): 2875-2887.

Elachi, C. (1997). Introduction to the physics and techniques of remote sensing. Wiley, New York, NY.

Emery, X (2004). Testing the Correctness of the Sequential Algorithm for Simulating Gaussian Random Fields. Stoch. Envir. Res. and Risk Ass. 18: 401-413.

Eumetsat (1998). Meteosat High Resolution Image Dissemination. Technical Description. Revision 4. Eumetsat Technical Report EUM TD 02. Darmstadt, Germany.

Eumetsat (2000). The Meteosat System, Revision 4. Eumetsat Technical Report EUM TD 05. Darmstadt, Germany.

Eumetsat (2004). "Meteosat VIS channel calibration information", http://www.eumetsat.de/en/dps/mpef/calibration/vis_calibration.html (Accessed on 5/12/2004).

Eumetsat (2006). "Meteosat Image Services > HRI - 0 degree", http://www.eumetsat.int/idcplg?IdcService=SS_GET_PAGE&nodeId=511&l=en (Accessed on March 2006).

Ferraro, R.R., Grody, N.C., Marks, G.F. (1994). Effects of Surface Conditions on Rain Identification Using the DMSP-SSM/I. Remote Sensing Reviews 11: 195–209.

Ferraro, R., Vincente, G., Ba, M., Gruber, A., Scofield, R., Li, Q., and Weldon, R. (1999). Satellite Techniques Yield Insight into Devastating Rainfall from Hurricane Mitch. Eos 80(43): 505-514.

Fiorucci, P., La Barbera, P., Lanza, L.G., and Minciardi, R. (2001). A geostatistical approach to multisensor rain field reconstruction and downscaling. *Hydrol. Earth Sys. Sci.* 5(2): 201-213.

Georgakakos, A.P. (2004). Decision Support Systems for integrated water resources management with an application to the Nile Basin. Key note Address, Proceedings, International Federation for Automatic Control Workshop on Modeling and Control for Participatory Planning and Managing Water Systems, Venice, Italy, Sep. 29 – Oct. 1.

Georgakakos, A.P., Yao, H., Brumbellow, K., De Marchi, C., Bourne, S., and Mullusky, M. (2000). Lake Victoria Decision Support System. Technical Report GWRI-2000-1. Georgia Water Resources Institute, Atlanta, GA.

Georgakakos, K.P., Tsintikidis, D., Attia, B., and Roskar, J. (2000). Estimation of Pixel-scale Daily Rainfall over Nile River Catchments Using Multi-Spectral METEOSAT Data. In Remote Sensing and Hydrology, 2000 (Proceedings of

IAHS symposium, Santa Fe, NM, April 2000), IAHS Publication No. 267, Owe, M., Brubaker, K., Ritchie, J., and Rango, A. (eds). Wallingford, UK.

Glacken, I.M., and Blackney, P.C.J. (1998). A Practitioner Implementation of Indicator Kriging. Presented at the "Beyond ordinary kriging" seminar of The Geostatistical Association of Australasia. Perth, Australia.

Gomez-Hernandez, J.J. (1991). A Stochastic Approach to the Simulation of Block Conductivity Fields Conditioned Upon Data Measured at a Smaller Scale. Ph.D. research, Department of Applied Earth Sciences, Standford University, Palo Alto, CA.

Grecu, M., and Anagnostou, E. N. (2001). Overland Precipitation Estimation from Passive Microwave Observations. J. Appl. Met. 40, 1367–1380.

Gremont, B. (2002) Simulation of Rainfield Attenuation for Satellite Communication Networks. 1st workshop of the COST 280 Action "Propagation Impairments Mitigation for Millimetre-Wave Radio Systems". Malvern, UK.

Griffith, C.G. (1987) Comparison of Et al. and Satellite Rain Estimates for the Central United States During August 1979. J. Geophysical Res. 92(D8): 9551-9566.

Griffith, C.G., Woodley, W.L., Browner, S., Teufeuri, J., Maier, M., Martin, D.W., Stout, J., and Sikdar, D.N. (1976). Rainfall Estimation from Geosynchronous Satellite Imagery During Daylight Hours. NOAA Technical Report. ERL 356 WMPO 7: Boulder, CO.

Hobbs, P.V. and Rango, A.L. (1985). Ice Particles Concentration in Clouds. J.Atmos. Sci. 42: 2523-2549.

Hossain, F., Anagnostou, E.N., and Dinku, T. (2004). Sensitivity Analyses of Satellite Rainfall Retrieval and Sampling Error on Flood Prediction Uncertainty. IEEE Transactions on Geoscience and Remote Sensing 42(1):130-139.

Houze, R.A. (1981). Structures of Atmospheric Precipitation Systems – a Global Survey. Radio Science 16: 671-689.

Houze, R.A. (1993). Cloud Dynamics. Academic Press, San Diego, CA.

Houze, R.A. (1997). Stratiform Precipitation in Regions of Convection: A Meteorological Paradox? Bulletin of the American Meteorological Society, 78(10): 2179-2196.

Houze, R.A. and Rappaport, E.N. (1984). Air Motions and Precipitation Structure of an Early Summer Squall Line of the Eastern Tropical Atlantic. J. Atmos. Sci. 41: 533 -574.

Hsu, K., Gao, X., Sorooshian, S., and Gupta, H.V. (1997) Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks. J. Appl. Met. 36: 1176-1190.

Hsu, K., Gupta, H.V., Gao, X., and Sorooshian, S. (1999). Estimation of Physical Variables from Multi- Channel Remotely Sensed Imagery Using a Neural Network: Application to Rainfall Estimation. Water Resources Research 35(5): 1605-1618.

Huffman, G.J., Adler, R.F., Morrissey, M.M., Bolvin, D.T., Curtis, S., Joyce, R., McGavock, B., ans Susskind, J. (2001). Global Precipitation at One-Degree Daily Resolution from Multisatellite Observations. J. Hydrometeorology 2(1): 36-50.

Iguchi, T. and Meneghini, R. (1994). Intercomparison of Single Frequency Methods for Retrieving a Vertical Rain Profile from Airborne and Spaceborne Data. Journal of Atmospheric and Oceanic Technology 11(6): 1507-1516.

Inoue, T. (1987). An Instantaneous Delineation of Convective Rainfall Areaas Using Split Window Data of NOAA-7 AVHRR. Journal of the Meteorological Society of Japan 65: 469-481.

Isaacs, R.G. (1993). Remote Sensing of Clouds for Defense and Climate Studies: an Overview. In Passive Infrared Remote Sensing of Clouds and the Atmosphere, D.K. Lynch, Editor, Proc. SPIE 1934, pp 2-35. SPIE, Washington, D.C.

Isaaks, E.H., and Srivastava, R.M. (1989). An introduction to Applied Geostatistics. Oxford University Press, New York, NY.

Johnson, D. (2006). “Satellite Coverages and Orbits”, http://www.rap.ucar.edu/~djohnson/satellite/coverage.html (Accessed on March 2006)

Johnson, L.R., Smith, P.L., Vonder Haar, T.H., and Reinke, D. (1994). The Relationship between Area–Time Integrals Determined from Satellite Infrared Data by Means of a Fixed-Threshold Approach and Convective Rainfall Volumes. Mon. Wea. Rev. 122(3): 440–448

Journel, A.G. (1989). Fundamentals of Geostatistics in Five Lessons. American Geophysical Union, Washington, D.C.

Joyce, R.J., Janowiak, J.E., Arkin, P.A., and Xie, P. (2004). CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. J. of Hydrometeorology 5: 487-503.

Kidd, C. (1998). On Rainfall Retrieval Using Polarization-Corrected Temperatures. International Journal of Remote Sensing 19: 981–996.

Kidd, C. (2001). Satellite Rainfall Climatology: A Review. International Journal of Climatology 21: 1041-1066.

Kidd, C., and Barrett, E.C. (1990). The Use of Passive Microwave Imagery in Rainfall Monitoring. Remote Sensing Review 4(2): 415-450.

Kidd, C., Kniveton, D.R., Todd, M.C., and Bellerby, T.J. (2003). Satellite Rainfall Estimation Using Combined Passive Microwave and Infrared Algorithms. J. of Hydrometeorology 4: 1088-1104.

King, P.W.S., Hogg, W.D., and Arkin, P.A. (1995). The Role of Visible Data in Improving Satellite Rain-rate Estimates. J. Appl. Meteor. 34: 1608-1621.

Kidder, S.Q. and Vonder Haar, T.H. (1995). Satellite meteorology: an introduction. Academic Press, San Diego, USA.

Kummerow, C., Barnes, W., Kozu, T., Shiue, J. and Simpson, J. (1998). The Tropical Rainfall Measuring Mission (TRMM) sensor package. J. Atmos. Oceanic Technol. 15: 809-817.

Kummerow, C., Hong, Y., Olson, W. S., Yang, S., Adler, R. F., McCollum, J., Ferraro R., Petty, G., Shin, D. B., and Wilheit, T. T. (2001). The Evolution of the

Goddard Profiling Algorithm (GPROF) for Rainfall Estimation from Passive Microwave Sensors. J. Appl. Met. 40, 1801–1819.

Kummerow, C., Simpson, J., Thiele, O., Barnes, W., Chang, A.T.C., Stocker, E., Adler, R.F., Hou, A., Kakar, R., Wentz, F., Ashcroft, P., Kozu, T., Hong, Y., Okamoto, K., Iguchi, T., Kuroiwa, H., Im, E., Haddad, Z., Huffman, G., Ferrier, B., Holson, W.S., Zipser, E., Smith, E.A., Wilheit, T.T., North, G., Krishnamurti, T. and Nakamura, K. (2000). The status of the Tropical Rainfall Measuring Mission (TRMM) after two years in orbit. J. Appl. Meteor. 39: 1965-192.

Kurino, T. (1997). A Satellite Infrared Technique for Estimating "Deep/Shallow" Precipitation. Adv. Space Res. 19(3): 511-514.

Lanza, L.G.(2000). A Conditional Simulation Model of Intermittent Rain Fields. Hydrology and Earth System Sciences 4(1): 173-183.

Li, Q., Bras, R. L. and Veneziano, D. (1996). Analysis of Darwin rainfall data: Implications on sampling strategies. J. Appl. Meteor. 35: 372-385.

Lopez, R.E., Thomas, J., Blanchard, D.O., and. Holle, R.L (1983). Estimation of rainfall over an extended region using only measurements of the area covered by radar echoes. Preprints, 21st Conf. Radar Meteor., Edmonton, Alberta, Amer. Meteor. Soc., 681-686.

Lovejoy, S., and Austin, G.L. (1979). The Delineation of Rain Areas from Visible and IR Satellite Data for GATE amd Mid-latitudes. Atmos. Ocen. 17: 77-92.

Mantoglou, A., and Wilson, J.L. (1982). The Turning Bands Method for Simulation of Random Fields Using Line Generation by a Spectral Method. Water Resour. Res. 18(5): 1379-1394.

Marsh, W.M. (1987). Earthscape: A Physical Geography. Wiley, New York, USA.

Marsh, W.M. and Dozier, J. (1986). Landscape: An Introduction to Physical Geography. Wiley, New York, USA.

Marshall, J.S. and Palmer, W.McK. (1948). The Distribution of Raindrops with Size. J. Meteorol. 5: 165-166.

Matheron, G. (1973). The Intrinsic Random Functions and their Applications. Advan. Appl. Prob. 5: 439-468.

McCollum, J. R., and Ferraro, R. R., (2003). Next Generation of NOAA/NESDIS TMI, SSM/I, and AMSR-E Microwave Land Rainfall Algorithms. J. Geophys. Res. 108, DOI:10.1029/2001JD001512.

Meneghini, R., Iguchi, T., Kozu, T., Liao, L., Okamoto, K., Jones, J.A. and Kwiatkowski, J. (2000). Use of the surface reference technique for path attenuation estimates from the TRMM precipitation radar. J. Appl. Meteor. 39: 2053 – 2070.

Menz, G. (1997). Regionalization of Precipitation Modes in East-Africa Using Meteosat Data. International Journal of Climatology, 17(10): 1011-1027.

Mitchell, T.D. and Jones, P.D. (2005). An Improved Method of Constructing a Database of Monthly Climate Observations and Associated High-resolution Grids. Int. J. of Climatology 25(6): 693-712.

Morrissey, M.L. and Janowiak, J.E. (1996). Sampling-induced conditional biases in satellite climate-scale rainfall estimates. J. Appl. Meteor. 35: 541 – 548.

NASA (2005). TRMM Senior Review Proposal 2005. National Aeronautics and Space Administration, Washington, D.C.

NASA- Goddard Space Flight Center (2006a). "TRMM Tropical Rainfall Measurement Mission", http://trmm.gsfc.nasa.gov/ (Accessed on April 2006).

NASA- Goddard Space Flight Center (2006b). "TRMM Tropical Rainfall Measurement Mission. Data Products and Descriptions", http://trmm.gsfc.nasa.gov/data_dir/ProductStatus.html (Accessed on April 2006).

Negri, A.J., Adler, R.F., and Xu, L. (2002). A TRMM-Calibrated Infrared Rainfall Algorithm Applied over Brazil. J. Geophys. Res. 107(D20): 8048.

Olson, W. S. (1989). Physical Retrieval of Rainfall Rates over the Ocean by Multispectral Microwave Radiometry: Application to Tropical Cyclones,' J. Geophys. Res. 94, 2267–2280.

Oz, B., Deutsch, C.V., Tran, T.T., and Xie, Y. (2003). DSSIM-HR: A FORTRAN 90 Program for Direct Sequential Simulation with Histogram Reproduction. Computers and Geosciences 29: 39-51.

Pardo-Iguzquiza, E., Grimes, D.I.F., and Teo, C. (2006). Assessing the Uncertainty Associated with Intermittent Rain Fields. Water Resour. Res. 42(1): 1412-1412.

Rigollier, C., Lefèvre, M., Blanc, Ph., and Wald, L. (2002). The operational calibration of images taken in the visible channel of the Meteosat-series of satellites. J. Atmos. Oceanic Technology 19(9): 1285-1293

Rinehart, R.E. (1991). Radar for Meteorologists, or, You Too Can Be a Radar Meteorologist. Rinehart, Grand Forks, N.D.

Richards, F. and Arkin P. (1981). On the Relationship Between Satellite-observed Cloud Cover and Precipitation. Monthly Weather Review, 105:446-457.

Scofield, R.A. (1987). The NESDIS Operational Convective Precipitation Technique. Mon. Wea. Rev. 115: 1773-1792.

Scofield, R.A., Oliver, V.J. (1977). A scheme for estimating convective rainfall from satellite imagery. NOAA Tech. Memo. NESS 86. NOAA/NESDIS, Washington, DC.

Seaspace Corp. (1999). Terascan Software Training Guide. Seaspace Corporation, San Diego, CA.

Seo, D.-J. (1998a). Real-time Estimation of Rainfall Fields Using Rain Gage Data Under Fractional Coverage Conditions. J. Hydrol. 208(1-2): 25-36.

Seo, D.-J. (1998b). Real-time Estimation of Rainfall Fields Using Radar Rainfall and Rain Gage Data. J. Hydrol. 208(1-2): 37-52.

Shahin, M. (1985). Hydrology of the Nile. Elsevier, Amsterdam, The Netherlands.

Smith, E.A. (2001). Satellites, Orbits and Coverages. Paper presented at IGARSS 2001 (International Geoscience and Remote Sensing Symposium), IEEE, Sydney, Australia.

Smith, E. A., Xiang, X., Mugnai, A. and Tripoli, G. J. (1994). Design of an Inversion-Based Precipitation Profile Retrieval Algorithm Using an Explicit Cloud Model for Initial Guess Microphysics. Meteor. Atmos. Phys. 54, 53–78.

International Conference on Interactive Information andSoares, A. (2001). Direct Sequential Simulation and Cosimulation. Math. Geology 33(8): 911-926

Steiner, M., and Houze, R.A. (1998). Sensitivity of Monthly Three-Dimensional Radar-Echo Characteristics to Sampling Frequency. Journal Of The Meteorological Society Of Japan 76(1): 73-95.

Steiner, M., Houze, R.A., and Yuter, S.E. (1995). Climatological Characterization of Three-dimensional Storm Structure from Operational Radar and Rain Et al. Data. J. Appl.Meteor. 34:1978-2007.

Thorne, V., Coakeley, P. Grimes, D. and Dugdale, G. (2001). Comparison of TAMSAT and CPC Rainfall Estimates with Rainet al.s, for Southern Africa. Int. J. Remote Sensing 22(10): 1951-1974

Todd, M.C., Barrett, E.C., Beaumont, M.J. and Green, J.L. (1995). Satellite identification of rain days over the Upper Nile river basin using an optimum infrared rain no-rain threshold temperature model. J. App. Meteor. 34(12): 2600-2611.

Todd, M.C., Barrett, E.C., Beaumont, M.J., and Bellerby, T.J. (1999) Estimation of Daily Rainfall Over the Upper Nile River Basin Using a Continuously Calibrated Satellite Infrared Technique. Meteorological Applications6(3): 201-210.

Todd, M.C., Kidd, C., Kniveton, D., and Bellerby, T. J. (2001). A Combined Satellite Infrared and Passive Microwave Technique for Estimation of Small-Scale Rainfall. J. Atmos. Oceanic Technol., 18, 742–755.

Tsonis, A.A., and Isaac, G.A. (1985) On a New Approach for Instantaneous Rain Area Delimitation in the Midlatitudes Using GOES Data. J. Clim. Appl. Meteorol. 24: 1208-1218.

Tsonis, A.A., G.N. Triantafyllou, and K.P. Georgakakos (1996). Hydrological Applications of Satellite Data. 1. Rainfall Estimation. J. Geophys. Res. 101(D21), 26517-26525.

Turpeinen, O.M., Abidi, A., And Belhouane, W. (1987). Determination of Rainfall with the ESOC Precipitation Index. Monthly Weather Review 115: 2699-2706.

University of Illinois (2002). "WW2010 Project - On Line Guides", http://ww2010.atmos.uiuc.edu/(Gh)/guides/home.rxml (Accessed on May 2002)

Vincente, G.A., Scofield, R.A. and Menzel, W.P. (1998). The operational GOES infrared rainfall estimation technique. Bull. Am. Met. Soc. 79: 1883-1898.

Walser, A., Lüthi, D., and Schär, C. (2004). Predictability of Precipitation in a Cloud-Resolving Model. Mon. Wea. Rev. 132: 560-577.

World Meteorological Organization (1981). Guide to Hydrological Practice, Vol. 1. Data Acquisition and Processing. World Meteorological Organization, Geneva, Switzerland.

Xu, L., Gao, X.. Sorooshian, S., Arkin, P.A., and Imam, B. (1999). A Microwave Infrared Threshold Technique to Improve the GOES Precipitation Index. J. Appl. Meteor. 38: 569-579.